Topological plasticity of a dual promoter gene: The *Drosophila melanogaster Adh* gene

by

Kaoru Miyahara

A thesis submitted for the degree of

Doctor of Philosophy

at

The Australian National University

March 1987

Statement

The data presented in this thesis were obtained entirely through my own work.

The work in this thesis has not been previously published by others except where due reference is made in the text. Some parts of the paper [Naora, H., Miyahara, K. & Braithwaite, A. W., (1984) J. Theor. Biol. **116**, 313] have been slightly modified to conform the thesis format.

K. Miyahara

Acknowledgements

I am greatly indebted to my supervisor Dr. Hiroto Naora for his stimulating discussion and encouragement throughout the course of this work. I would like to thank for Dr. William P. Diver and Dr. Antony W. Braithwaite for their helpful discussion.

I am also very grateful to all members in our group for their friendship and kindness.

My special thanks go to Ms. Honami Naora and Ms. Kyoko Koishi for typing this thesis and to Mr. David Buckle for correcting my English in this thesis.

I wish to thank the Australian National University for their Ph.D. Scholarship.

ABBREVIATIONS

Adh	alcohol dehydrogenase (gene)
ADH	alcohol dehydrogenase (protein)
bp	base pairs
c-onc	cellular oncogene
DNase I	deoxyribonuclease I
EDTA	ethylenediaminetetraacetic acid
EGTA	ethyleneglycol-bis-(b-aminoethyl ether)-N,N,N′,N′-tetraacetic acid
endo VII	endonuclease VII
Hepes	N-2-Hydroxyethylpiperane-N′-2-ethanesulfonic acid
kb	kilobase pairs
LTR	long terminal repeat
MBNase	mung bean nuclease
PDE I	phosphodiesterase I
SDS	soduim dodecyl sulfate
SSB	single-strand binding protein
SSC	saline soduim citrate [20xSSC contains 0.3M citric buffer (pH 7.0) and 3M NaCl]
Tris-Cl	Tris(hydroxymethyl) aminomethane - HCl

Other abbreviations are defined in the text.

ABSTRACT

A gene has extragenic territorial DNA sequences required for active function. The main aim of my thesis is to provide information on the molecular basis of gene territories. Using the *Drosophila melanogaster* alcohol dehydrogenase (*Adh*) gene, a typical dual promoter gene, I have attempted several lines of experiments and made the following observations.

1) It was found the DNA molecule containing the *D. melanogaster Adh* gene exhibits a particular plasticity of structure. For example, using various nucleases and reagents, it was found that several kinds of altered DNA structure are induced at the *Adh* locus by torsional stress. These structures can interact in a dynamic way according to torsional stress and ionic conditions, probably because they compete with one another for torsional energy in a superhelical domain. Such dynamic interaction may take place over a long distance in a superhelical domain without direct physical contact. Some altered DNA structures are located at functional sites of the *Adh* gene, e.g., two capping sites and poly(A) addition site.

2) I have found that the distribution of thermodynamic stability of the DNA duplex along the *Adh* gene is correlated with the functional organization of the *Adh* gene. The pattern of distribution observed at the *D. mulleri Adh* locus is the same to that at the *D. melanogaster Adh* locus. These observations suggest that the DNA molecule of eukaryotic genes may be spatially organized into various functional parts with specific structural properties, and conserved during evolution.

3) I have analyzed the DNase I hypersesitive sites in chromatin of the *Adh* gene at various stages during development. The distribution pattern of the DNase I hypersensitve sites alters on a stage-specific manner in concert with the transcriptional utilization of dual promoters of the *Adh* gene. This concerted alteration strongly suggests that specific conformation of chromatin structure is requisite for a specific promoter of the transcriptional unit to be accessible to transcriptional trans-acting factors.

4) Some DNaseI hypersensitive sites on chromatin are located at the positions corresponding to the altered DNA structures on supercoiled DNA. It is suggested that secondary structures or altered DNA structures are recognizable in distinct protein-DNA interactions, and play an important role in construction of the higher order structure of active chromatin under torsional stress.

5) The *Adh* gene appears to be demarcated by both altered DNA structures and the DNase I hypersensitive structures at the specific positions corresponding to the borders of intrinsic territories as predicted.

6) I have discovered a neighboring gene upstream of the *Adh* gene as predicted on the basis of the gene territoriality hypothesis. In conclusion, gene territoriality probably plays a crucial role in coordinated regulation of clustered genes through conformation of higher order structure.

CONTENTS

C H A P T E R O N E

GENERAL INTRODUCTION

1.1 Gene Territoriality

Genes of higher eukaryotes are well dispersed on chromosomes. The genes are not contiguous (as are prokaryotic genes) but are interspersed with non-functional DNA sequences (Graham et al., 1974). A question arises as to whether there exists a complex interaction between "non-functional" DNA sequences and functional genes regarding the observed spatial arrangement of eukaryotic genes.

A huge excess of "non-functional" DNA is present in the eukaryote cell nucleus. This is often referred to as the 'C-value paradox' (reviewed by Cavalier-Smith, 1985). Only a small portion of the genome of eukaryotes is transcribed at each stage in their life (Davidson, 1976). A great majority of single-copy DNA sequences do not appear to serve any selectively advantageous function and are hence considered dispensable. However, it should be noted that these DNA sequences together with repetitive DNA sequences appear to have been retained on chromosomes over a long evolutionary period. The results of computer simulation experiments suggest that random duplication and deletion events are capable of generating a complex genome carrying a large amount of dispensable sequences as vestiges (Loomis & Gilgin, 1986). A recent hypothesis (Naora et al., submitted) suggests that a primordial genome structure (composed of a 0.55 kb long reading frame and 19.5 kb long non-coding adjacent sequence) is retained in present-day chromosomes even after a series of gene

duplication events.

The concept of gene-to-gene interactions first arose as a result of extensive studies on the spatial arrangements of various clustered genes on eukaryotic chromosomes (Naora & Deacon, 1982a). That analysis suggested that the intergenic distance between a given pair of genes was a function of the combined gene size. (see Fig. 3.3.1) The extragenic DNA sequences were referred to as 'gene territories'. In general, it appears that the intergenic distance increases with increasing gene size. Such a gene size-dependent increase of the intergenic distance is, however, not seen if the total size of a pair of genes is less than 0.3 kb or greater than 5 kb. Therefore, there seems to be a minimum (around 0.3 kb long) and a maximum (around 13.5 kb long) size for the intergenic region between a pair of genes. From these observations, it has been proposed that genes on eukaryotic chromosomes are surrounded by extragenic territorial DNA sequences or non-functional DNA sequences which ostensibly serve to maintain the proper function of the genes. Furthermore, the authors also noted that when two genes lay on the same DNA strand (i.e. had the same direction of transcription) and had an intergenic distance shorter than a defined length, the transcription of one or both genes was inactivated or reduced, a phenomenon termed the 'territorial effect' (Naora & Deacon, 1982b). It has been suggested that this effect is analogous to ecological dynamics in that it mimics the associated, interactive behaviour of living organisms with their environment. On the basis of this, it

was proposed that an 'ecological' interaction may exist between genes at the molecular level.

The extragenic territorial DNA sequences required for gene activity may presumably confer a selective advantage. This may have an evolutionary significance - being related to the stable maintenance of the proper function of duplicated genes and the coordinated regulation of gene expression. The detailed molecular mechanisms by which these extragenic territorial DNA sequences may function is currently unknown. However, it has been suggested that territorial DNA sequences may contribute in part to the formation of a higher order chromatin structure which is necessary for efficient transcriptional activity and its regulation (Naora & Deacon, 1982a).

There are several cases in which extragenic territorial DNA sequences may play a key role in gene regulation (Naora & Deacon, 1982b; Naora et al., 1983). Transposable elements and retroviruses are known to alter the expression of neighbouring genes by their transposition into host chromosomes. For example, the activation of cellular chromosomal genes, e.g. *c-myc* (cellular homologue to the viral *myc* oncogene), appears to result from the insertion of a strong promoter present in long terminal repeats of proviruses into the vicinity (Hayward et al., 1981). However, examination of the location of proviruses in avian leukosis virus (ALV)- induced bursal lymphomas has revealed that an activation of *c-myc* oncogene is associated

with at least three different configurations of ALV: ALV was inserted (1) to the 5'-side of *c-myc* in the same transcriptional orientation; (2) to the 5'-side, but in the opposite orientation, or (3) to the 3'-side of the *c-myc* gene in the same orientation (Payne et al., 1982). Therefore, transcriptional activation of neighbouring genes is not due solely to viral promoter insertion. It seems likely that interaction between the extragenic territories of cellular genes and transposed proviral genes is involved in those cases.

The *gypsy* transposable element can cause mutations in the suppressible *forked*, *yellow* or *bithorax* alleles of *Drosophila* (Parkhurst & Corces, 1985; Modolell et al., 1983; Bender et al;., 1983). In particular, mutations at the unlinked suppressor of the *Hairy-wing* locus result in an alteration of transcription of these genes indirectly through interaction with *gypsy*. It was proposed that the mutational effect of the *gypsy* element may be due to interference with the *yellow* locus at a specific developmental stage (Parkhurst & Corces, 1986). The phenotypes and DNA lesions of *abx* and *bx* alleles in the mutants have been characterised and the results were interesting: the strength of their phenotypes varies in a regular manner dependent on the position and orientation of the *gypsy* element inserted (Peifer & Bender, 1986). These observations suggest that mutual interactions between these genes and the *gypsy* gene play an important role in their

This suggestion is strongly supported by recent observations (R.Hill, personal communication).

regulation. It is likely that an interaction between gene territories may play a part in this phenomenon.

Insertional mutagenesis can be experimentally achieved by introducing the Moloney leukemia provirus (~~M-MuLV~~ M-MoMLV) into a mouse germ line (Jaenish et al., 1981; 1983). A proviral insertion interfered with the synthesis of stable α1(I) collagen RNA (Schnieke et al., 1983) and caused a recessive lethal mutation. An analysis of DNase I hypersensitive sites on the mutant DNA revealed that the chromatin structure of the collagen gene was markedly altered. It has been suggested that the proviral insertion may prevent regulatory changes in the DNase I hypersensitive structures, thereby interfering with the proper activation of the collagen gene during development (Breindl et al.,1984). As in previous cases, such a proviral insertion would cause a serious interaction between gene territories and thus this interference could possibly be explained in part by 'territorial effects'.

Recently Emerman and Temin (1982) have attempted to demonstrate gene-to-gene interactions in reconstructed retroviruses. They infected cells with a reconstructed retrovirus containing two genes, which existed in the same transcriptional orientation with a very short intergenic distance. In those infected cells containing a single copy of the proviral form of the retrovirus, the expression of one gene was suppressed whenever the other gene was selectively expressed. It appears that both genes cannot be expressed simultaneously. The amount of suppression, in

fact, correlated with a decrease in the amount of transcribed RNA (Emerman & Temin, 1986a). It should be noted that this suppression is cis-acting, epigenetic and reversible and in these respects, very similar to the proposed territorial effects of genes.

A similar observation was made using a reconstructed Moloney leukemia virus containing the rabbit β-globin gene and the neomycin gene (Jaenish et al., 1986). In the presence of neomycin, the β-globin gene was not expressed especially in cells transformed with viruses containing both genes adjacent to each other and lying in the same transcriptional orientation. In contrast, the β-globin gene was actively expressed in cells which were transformed with viruses containing both genes in opposite orientation. These observations are in accordance with those predicted by the gene territoriality hypothesis, i.e. the territorial effects of genes may be brought to bear on each other only when the genes lie in the same transcriptional orientation, as previously mentioned. It is possible, for instance, that a mutually exclusive promoter choice may correspond to an assembly of mutually exclusive chromatin structures. Assumably territorial effects may be brought to bear through such mechanisms.

Almost all clustered genes in higher eukaryotes are expressed in a tissue specific manner or, alternatively in a stage-specific manner throughout development. It is

possible that in such cases the relationship between intergenic distance and total gene size may serve as a mechanism for the regulation of gene expression. In the case of closely linked genes with an intergenic distance shorter than the defined length, however, the genes never appear to be expressed simultaneously. For example, the rat α- and β- myosin heavy chain (MHC) genes lie close together in the same transcriptional orientation (Mahdavi et al., 1984), and may be considered to be under the influence of territoriality. During heart development, the expression of the β-MHC gene is predominant in late foetal life, but is completely replaced post-natally by α-MHC gene expression (Lompre et al., 1984). Both genes are also alternatively expressed in response to hormonal and pathological stimuli, i.e. when the α-MHC gene is activated, the β-MHC gene becomes inactivated and *vice versa*. Alternate expression of genes is also apparent in various other cases, e.g. the chicken β and β^H globin genes (Dolan et al., 1971; Londes et al., 1982), the ovalbumin and X genes (Lawson et al., 1980) and the δ_1 and δ_2 lens crystallin genes (Howkins et al., 1984; Borras et al., 1985). It should be mentioned here that all of the genes showing such alternate expression may be considered to be under territorial confrontation. Therefore, the arrangement of these genes (i.e. the resulting territorial DNA sequences) may provide a secure control mechanism for gene expression networks during development.

All these observations suggest that mechanisms in which territorial effects among neighbouring genes are involved may operate widely to regulate gene expression in eukaryotic cells. At present the molecular mechanism(s) underlying these territorial effects are not known, but some possible molecular mechanisms are proposed below.

1.2. Topological Plasticity

The chromosome is variable in shape due to the ability to transform DNA molecules into a condensed structure. The chromosome of *E. coli* is composed of a single circular double-stranded DNA molecule. This circular chromosome can be transformed into many loops held together by RNA and proteins (Stonington & Pettijohn, 1971). Many viral chromosomes also form a circular structure. However, the molecular structures of eukaryotic DNA present on chromosomes are much more complex than those of *E. coli* and viral DNAs. The eukaryotic DNA molecules are separately organised in respective chromosomes, i.e. X- and Y-chromosomes and several autosomes. Although the circularity of eukaryotic chromosomal DNA has often been questioned, it is widely accepted that the chromosomal DNA molecules are partitioned into many domains via loop structures similar to those found in *E. coli* (Pienta & Coffey, 1984). The size of these loops appear to range from around 20 - 200 kb (Paulson & Laemmli, 1977; Benyajati & Worcel, 1976). A loop structure is fixed at a point to the nuclear matrix, but the fixed points are considered to be dynamically changeable during cell cycles and developmental stages (Vogelstein et al., 1980; Buegiono-Nardelli et al., 1982). From the topological point of view, therefore, a closed loop structure in the eukaryotic nucleus is considered to be equivalent to that of a closed DNA circle. Thus, this structure is a universal one, being common to all DNA of biological origin. For this reason, data on topological conformation and transcriptional

activity of closed circular DNA, e.g. viral and recombinant plasmid DNAs, will provide us with valuable information about the structure and transcriptional regulation mechanisms within closed loop structures present on eukaryotic chromosomes.

Circular double-stranded DNA molecules isolated from both prokaryotic and eukaryotic cells are negatively supercoiled. There are two different mechanisms by which supercoils may be introduced into DNA molecules. In prokaryotes, negative supercoils are introduced enzymatically by DNA gyrase in an ATP-dependent manner (Gellert et al., 1976). Negative supercoil in eukaryotes appears to be introduced by the coiling of duplex DNA around nucleosomes mediated by a nicking-closing enzyme (Germond et al., 1975; Laskey et al., 1977). All superhelical turns in eukaryotic DNA were once thought to be restrained in nucleosomes (Sinden et al., 1980). However, recent experiments have suggested that eukaryotic DNA in chromatin is also under torsional stress.

It is well known that torsional stress in DNA molecules can lead to the formation of altered DNA structures at certain DNA sequences. Some of the altered DNA structures can be detected by digestion with S1 nuclease or chemical reagents. Specifically, cruciform structures (Lilley, 1980; Panayotatos & Wells, 1981) and Z-form structures (Singleton et al., 1982) are apparently inducible by torsional stress at inverted repeats and at alternating purine-pyrimidine sequences respectively.

Other DNA sequences, e.g. A-T rich regions (Panayatatos & Wells, 1981), short direct repeats (Mace et al., 1983) and homopurine-homopyrimidine tracts (Schon et al., 1983), are also detectable by S1 nuclease digestion only when the molecules are under torsional stress. The structural nature of these S1 nuclease sensitive sites is unknown. It is apparent, however, that DNA conformational alterations from B-form to non-B-form structures are induced at these specific sequences by torsional stress.

Using a sensitive S1 nuclease detection method, the location of such altered DNA structures on recombinant DNAs has been determined. Sites hypersensitive to S1 nuclease are often detected upstream or downstream from the protein coding regions of various eukaryotic genes (Larsen & Weintraub, 1982; Selleck et al., 1984). Similarly, DNase I hypersensitive sites in chromatin are also observed frequently at or near the 5′- and 3′- ends of genes (reviewed by Eissenberg et al., 1985). The appearance of DNase I hypersensitive sites is often associated with the induction of or a commitment to gene expression. Of particular interest is that some of the S1 nuclease hypersensitive sites in naked DNA molecules under torsional stress are also sensitive to DNase I in chromatin *in vivo* (Larsen & Weintraub, 1982; Han et al., 1984; Jordano & Perucho, 1986). Therefore, it seems likely that a conformational alteration of DNA molecules at specific sites plays an important role in the formation of DNase I hypersensitive structures in chromatin during transcriptional activation of a gene.

Recently, experiments designed to monitor the requirement for topological alteration during transcription were carried out. Circular recombinant DNAs containing a gene can be transcribed much more efficiently than linear DNA following microinjection into *Xenopus* oocyte nuclei (Melton et al., 1983; Probst et al., 1979). Harland et al (1983) showed that linearisation of template DNA reduced its transcriptional activity. This experiment, however, did not show whether more active transcription correlated with closed circularity or supercoiling of the template DNA. Subsequently, it has been reported that the addition of a gyrase inhibitor (novobiocin) and DNase I causes the relaxation of half the minichromosome of injected recombinant DNA and a concomitant drastic reduction of template activity (Ryoji & Worcel, 1984). Similarly, it has been found that only a transcriptionally active fraction of SV40 minichromosomes separated in sucrose gradient possesses a DNase I hypersensitive region and can be relaxed by topoisomerase I (Luchnik et al., 1982, 1985). Moreover, an interesting observation in intact cells is that novobiocin can induce a reversal in active chromatin structure in chicken red blood cells, so that the β-globin gene region is no longer sensitive to DNase I (Villeponteau et al., 1984). It has also been shown that DNA topoisomerase I of *Drosophila* is preferentially associated with transcriptionally active loci and that the distribution of this enzyme displays a pattern similar to that of RNA polymerase II (Fleischmann et al., 1984;

Gilmour et al., 1986). All these observations strongly suggest that a torsionally strained chromosome may correspond with transcriptionally active chromatin. The state of torsional stress may be regulated by the action of both types of topoisomerases.

From the foregoing, I believe that an investigation of the conformational alteration of DNA molecules and the higher order structure of chromatin will contribute to a better understanding of the regulatory mechanisms underlying the dynamics of gene expression.

1.3 A dual promoter gene; the *Drosophila melanogaster* alcohol dehydrogenase gene

In eukaryotes, RNA polymerase II transcribes DNA sequences which often encode a single transcriptional unit, and hence a single species of mRNA is produced. A transcriptional unit is equipped with the signals necessary for initiation and termination of transcription at the appropriate sites. However, the production of multiple mRNA species from a single gene locus is now well documented. These gene transcripts possess alternative signals for various steps of mRNA production and processing (reviewed by Leff et al., 1986). Most of such complex transcriptional units in a gene locus appear to encode multiple species of proteins via their respective mRNAs. However, some units may still encode a single species of protein. In the latter case, multiple mRNAs from transcriptional units in a single gene locus contain identical protein coding sequences, but different sequences at the 5′- and/or 3′- untranslated regions. Several genes with dual promoters or alternative 5′- initiation sites have also been reported. They include the mouse α-amylase gene (Young et al., 1981; Shaw et al., 1985), the human *c-myc* gene (Battey et al., 1983), and the *D. melanogaster* alcohol dehydrodenase (*Adh*) gene (Benyajati et al., 1983).

The *Adh* gene has been extensively characterised. In *D. melanogaster*, the *Adh* gene is expressed in a tissue-specific manner and its activity is regulated during

development. The ADH enzyme activity is detected predominantly in the fat body and mid-gut of the larvae and adult fly (Ursprung et al., 1970). The *Adh* gene is transcribed into two different mRNAs using distinct promoters, proximal and distal, which are 708 bp apart (Benyajati et al., 1983). The pre-mRNA transcript from the distal promoter contains a 654 base long 5′- intron sequence that overlaps with the 5′- end of the pre-mRNA transcript from the proximal promoter. Both mRNA transcripts possess identical protein coding sequences and identical 3′- ends. Only the proximal transcript is produced during the stage from late embryogenesis to mid-third larval instar, and thereafter the level of this transcript falls rapidly. Concomitantly, the distal transcript accumulates, the level increasing rapidly at eclosion. The distal transcript is also predominant in an adult fly (Savakis et al., 1986).

In experiments using P element-mediated transformation, the cis-acting sequences necessary for both tissue specificity and developmental expression of the *Adh* gene have been shown to reside in an 11.8 kb genomic DNA fragment (Goldberg, 1980). Furthermore, cis-acting control elements and trans-acting factors for transcription of the *Adh* gene have been analysed in *in vitro* transcription systems (Heberlein et al., 1985). The analysis suggests that multiple sequence-specific DNA binding proteins may interact differentially with the proximal and distal promoters of the *Adh* gene for transcriptional activation.

However, the endogenous *Adh* gene in cultured cells, which appears to have all the necessary transcriptional machinery, is transcriptionally inactive (Benyajati & Dray, 1984). It would appear that the endogenous gene is assembled in an inactive chromatin configuration which does not permit recognition of the promoters or other cis-acting elements by trans-acting factors. Thus, such trans-acting factors may be necessary, but are apparently not sufficient for transcriptional regulation of the *Adh* gene. We have, as yet, no information about the chromatin configuration of the *Adh* gene. The molecular mechanisms underlying regulation of these dual promoters remain unsolved.

As mentioned previously (Chapter 1.1), an individual gene requires a gene size-dependent length of extragenic DNA sequences for active function. The gene size has been defined as a transcriptional unit, starting from the capping or homologous 5′- terminal site and ending at the sequence which corresponds to the poly(A) addition or homologous 3′- terminal site. This means that in the case of a dual promoter gene, the gene size is variable depending upon which of the two promoter is utilised for transcription. In the *Adh* gene, the size of the distal transcriptional unit is about 1.8 kb and that of the proximal unit is about 1.1 kb. According to the gene territoriality hypothesis, these alternative units may be estimated to possess about 2.9 kb and 0.9 kb long intrinsic territories, respectively, at both the 5′- and 3′- flanking regions (see Chapter 3.1). Therefore, questions arise as

to (1) whether a dual promoter gene requires multiple lengths of extragenic DNA sequences; and (2) whether such a gene is organised in various higher order structures of chromatin in a manner that corresponds to the usage of specific promoters.

1.4 Hypothesis and aims

As described previously, an individual gene should have its own extragenic territorial DNA sequences for active function (Chapter 1.1). The main aim of my thesis is to provide information on the molecular basis of gene territoriality and also to confirm certain predictions derived from the gene territoriality hypothesis. A simple prediction from this hypothesis is that genes which possess multiple initiation sites and thereby differing in the length of transcriptional units, should have multiple extragenic 'territorial' regions differing in length according to the length of these respective transcriptional units. Does a dual promoter gene, such as the *D. melanogaster Adh* gene, in fact have two alternative territories as predicted? If so, what are the basic properties of these territories at the molecular level? As the *D. melanogaster Adh* gene, as mentioned previously (Chapter 1.3), displays characteristic features not only in terms of its structure but also in the regulation of its expression, it is an ideal candidate gene for resolving these questions in detail. It is for this reason that an attempt was made to characterise the proposed "dual" territories of the *Adh* gene. Particular efforts were made to investigate the following topics:

(1) Topological plasticity of the DNA molecule around the *Adh* gene

Dynamic aspects of basic properties of gene territoriality are examined in Chapter 3.1. This study included an examination of the possibility that a gene may have intrinsic territories in the 5′- and 3′- flanking regions and that the demarcation points of these intrinsic gene territories may be marked by specific conformational properties. It is likely that DNA molecules do not exhibit a rigid B-form structure in solution, but can form different structures at particular sites depending upon the microenvironment and torsional stress on the molecule. It is for this reason that the alterations of topology around the *Adh* gene were investigated in detail under various conditions (see Chapter 3.2 and 3.3).

(2) Torsional stress and transcriptional activity

The plasticity of a DNA molecule is a result of torsional stress. It is likely that this plasticity plays an important role in gene function. Indeed, it has been revealed that torsional stress has effects on transcriptional activity (Chapter 3.4). This study was based on the possibility that the topological plasticity of DNA molecules under torsional stress may relate to transcriptional control.

(3) Studies on higher order structures of chromatin

The alteration of chromatin higher order structure is associated with transcriptional activity. Does higher order structure change concomitantly with the switch of transcription between the dual promoters of the *Adh* gene? Furthermore, does the plasticity of DNA molecules directly participate in the configuration of higher order structures of the *Adh* gene? To answer these questions, the higher order structures of chromatin on the *Adh* gene during development were investigated by digestion with DNase I (Chapter 3.5).

(4) Survey of neighbouring genes, as predicted by the gene territoriality hypothesis

Expressions of oncogenes and other genes are altered by chromosomal translocation and virus integration. Many cases of such alterations have been considered in the light of gene territoriality, which implies that neighbouring genes should exist around these genes. An important prediction based on the gene territoriality hypothesis is that the expression of a gene and its neighbours affect one anotherand are mutually regulated (Chapter 3.6). Two transcriptional units in the *Adh* gene are alternately expressed during development. Based on the gene territoriality hypothesis, I predicted the presence of a neighbouring gene(s) which is/are involved in the

regulation of expression of these transcriptional units. An attempt was made to confirm this prediction (Chapter 3.7).

C H A P T E R T W O

GENERAL MATERIALS AND METHODS

2.1. Restriction of the DNA molecule

Restriction endonucleases were purchased from Boehringer-Mannheim or New England Biolabs. DNA molecules were digested by incubation with a restriction endonuclease in Hae III buffer [10 mM Tris-Cl (pH 7.5), 6 mM $MgCl_2$, 10 mM 2-mercaptoethanol and 100 μg/ml gelatin] containing an appropriate concentration of NaCl or KCl according to the manufacturers' instructions unless otherwise stated in respective experiments.

2.2. Plasmid DNA preparation

E. coli K12 strains HB101 and JM107 were used as host cells. A single colony of transformed HB101 cells were grown in 500 ml LB medium (Maniatis et al., 1982) in the presence of 50 μg ampicilin per milliliter at 37°C with vigorous shaking. When the culture reached an OD_{650} of 0.5-0.7, 170 μg of chloramphenicol per milliliter were added for amplification, and then incubation was continued for longer than 12 hrs. In the case of JM107 cells, the amplification by chloramphenicol was omitted. Supercoiled plasmid DNAs were prepared according to Triton lysis procedures (Pratt, 1984). Occasionaly the DNAs were purified by CsCl gradient twice. Purified DNAs were stored in TE buffer [10 mM Tris-Cl (pH 7.5), 1 mM EDTA] at -20°C.

2.3. Southern hybridization

Restriction enzyme-treated DNAs were electrophoresed on 1% (w/v) agarose gel in 1 x TAE buffer [40 mM Tris-acetate (pH 8.1), 2 mM EDTA]. The gel was treated with 1.5 M NaCl-0.5 M NaOH and then 3 M NaCl-0.5 M Tris-Cl (pH 7.4), and blotted onto nitrocellulose filter (Schleicher & Schuell Co Ltd.) with 20 x SSC for longer than 12 hrs (Southern, 1975). The blotted filter was baked at 80°C for 2 hrs under vacuum and hybridized at 65°C in Southern hybridization buffer [50 mM Hepes (pH 7.0), 10 x Denhardt's solution, 3 x SSC, 0.1% SDS, 20 μg/ml t-RNA, 20 μg/ml denatured salmon sperm DNA]. The hybridized filter was soaked in 2 x SSC - 0.1% SDS at room temperature, in the same solution at 65°C, and in 0.1 x SSC - 0.1% SDS at 65°C twice, sequentially. The filter was exposed on an X-ray film (XAR-5, Kodak) at -80°C for an appropriate time with or whithout an intensifying screen.

2.4. Making the hybridization probe

A DNA fragment was labelled with [α-^{32}P] dATP by random priming procedures (Feinberg & Vogelstein, 1983). A DNA fragment cloned into Bluescribe vector M13 (Vector Cloning Systems Co., Ltd.) was transcribed to produce high specific activity strand-specific RNA probes according to the directions of the company.

The typical specific activity of probes was 10^8-10^9cpm per microgram DNA.

2.5. Fragment isolation

Restricted DNA fragments were run on 1% agarose gel, then an appropriate DNA fragment was electro-eluted onto a piece of dialysis membrane (Yang et al., 1979). The DNA was purified with PCIA [phenol:chloroform:isoamylalcohol 50:49:1 (v/v)], and precipitated with ethanol, then stored in TE buffer at -20°C. In the case of DNA fragments for *in vitro* transcription, the DNA was purified through DEAE-Sephacel (Pharmacia) (Maniatis et al., 1982).

2.6 Preparation of superhelical topoisomers

Supercoiled closed circular DNA was purified by two times centrifugation to equilibrium in CsCl-ethidium bromide gradient. Ethidium bromide was completely extracted from plasmid preparation with water-saturated n-butanol. The purified DNA was incubated in 100 µl of a solution of 50 mM Tris-Cl (pH 7.6), 10 mM $MgCl_2$, 0.1 mM dithiothreitol, 2 mM spermidine, with various concentrations of ethidium bromide (0-100 µM) and topoisomerase I (calf thymus; BRL) for 12 hrs at 30°C in the dark (Singleton & Wells, 1982a). The concentration of ethidium bromide was determined by measurement of optical densties at 460 and 287 nm. At 460 nm the specific absorptivity is taken as 4220 M^{-1} cm^{-1} and at 287 nm as 5.39×10^4 M^{-1} cm^{-1} (W. Keller, 1975). The treated DNA was purified by phenol/chloroform twice and precipitated with ethanol. The DNA topoisomer was stored in TE buffer at -

20°C. The mean negative superhelical density (σ) of DNA topoisomer was determined by electrophoresis in TBE buffer [80 mM Tris-phosphate (pH 8.3), 8 mM EDTA] containing a varied concentration of chroloquine (Shure et al., 1977). The ladders of topoisomers on the gel were hard to count due to the length of the DNA. Assumed a physiologcally supercoiled DNA has a superhelical density ($-\sigma$) of 0.055, the mean superhelical densities were estimated from the pattern of topoisomers on the gel.

C H A P T E R T H R E E

THEORETICAL CONSIDERATIONS

AND

EXPERIMENTS

3.1 Theoretical study of a gene-to-gene interaction: Dynamic aspects of the territories of a gene required for active expression

It has been widely accepted that various regulatory elements (e.g. TATA, CAAT and enhancer sequences) on eukaryote chromosomes are involved in the regulation of gene expression (Gruss, 1984; Yaniv, 1984). Actually, regulation of gene expression may result from various types of direct and indirect interactions with the transcriptional machinery (Lima-de-Faria, 1983; Croce et al., 1984; Webster & Goodwin, 1984; Weintraub, 1985). Recently, it has been shown that the spatial arrangement of genes on chromosomes appears to be involved in the regulation of gene expression in higher eukaryotes (Naora & Deacon, 1982a; Naora, 1986).

It has been shown that clustered genes on eukaryote (mainly higher eukaryote) chromosomes are surrounded by extragenic DNA sequences which must be of a certain size for activity (Naora & Deacon, 1982a). The surrounding extragenic DNA sequences are called "gene territories". An interesting observation was that the intergenic distance, composed of extragenic territorial DNA sequences , between any two given genes in a cluster on the same DNA strand is defined by the sizes of the genes involved (Naora & Deacon, 1982a). When two genes lie on the same DNA strand and have an intergenic distance shorter than a defined length, the transcriptional activity of one or both genes is reduced or ceases in most cases. These we called "territorial

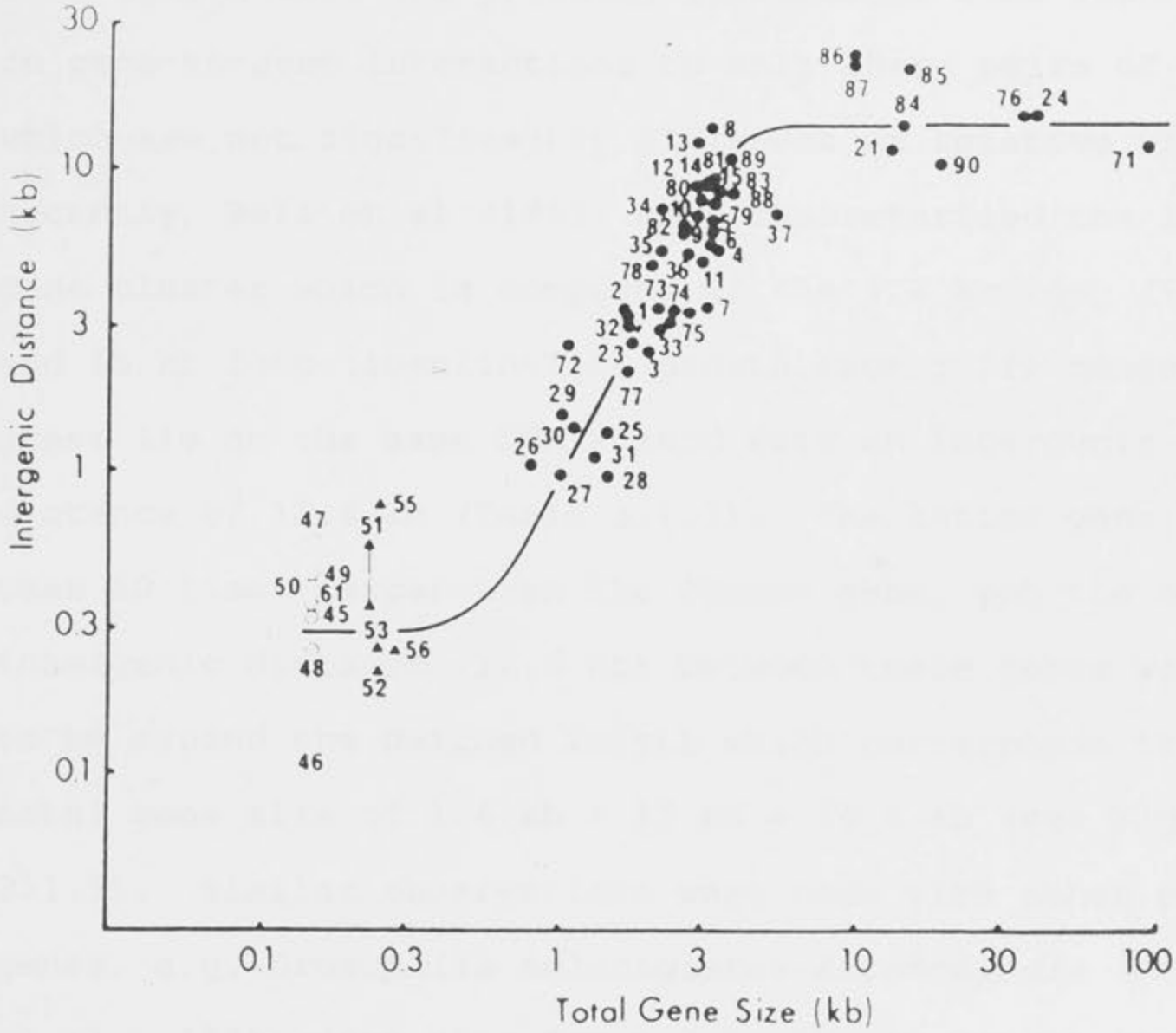
effects" (Naora & Deacon, 1982a and b). Recently, a similar observation was made by Emerman and Temin (1984) using reconstructed retroviruses. All of these observations indicate that the gene's activity is regulated by a cis-acting gene-to-gene interaction and implies that the size of a gene's territory is the key factor which determines the potential activity of clustered genes on chromosomes.

At present, we have little knowledge of the nature of these extragenic territories. In this chapter, an attempt was made to elucidate the extent to which the territory of a given gene is defined by a neighbouring gene within a cluster. I show here that the intrinsic territory is the specific feature of a given gene, but the actual territory is a function that involves a complex manner of interaction between genes.

It has been postulated that a gene cluster was formed mainly by gene duplication, followed by subsequent alterations, (i.e. base substitutions, deletions and insertions) and correction of gene's sequences (Ohno, 1970; Jeffreys, 1982). The sizes of most of these duplicated gene copies, however, have not changed significantly over long evolutionary periods. For example, despite the fact that the human σ- and β-globin genes genes diverged about 40 to 70 million years ago (Jeffreys, 1982) the difference in their size amounts to only 2% (Lawn et al., 1980; Spritz et al., 1980). The previous observation, which showed that the intergenic distance of a chosen pair of clustered genes

Fig. 3.1.1 Relationship between intergenic distances of gene pairs and total gene sizes (reproduced from Naora, 1986).

The logarithm of intergenic distances of the chosen gene pairs are plotted against the logarithm of total gene sizes of the paired genes. The points numbered represent various gene pairs cited in Naora and Deacon (1982a) and Naora (1986). For details, see above references.


30
10
3
1
0.3
0.1
Intergenic Distance (kb)
0.1
0.3
1
3
10
30
100
Total Gene Size (kb)

is defined by their total gene sizes was ~~maed~~ made by analysis of a large number of pairs of almost equally sized genes (see Fig. 3.1.1). It was for this reason that the previous discussions were restricted on gene-to-gene interactions to only these pairs of genes which are not significantly different in relative size. Recently, Bell et al (1985) have characterised the insulin gene cluster which is composed of the 1.4 kb long (insulin) and 15 kb long (insulin-like growth factor II) genes. Both genes lie on the same DNA strand with an intergenic distance of 12.6 kb (Table 3.1.1). The latter gene is more than 10 times larger then the former gene, yet the measured intergenic distance (12.6 kb) between these genes was found to be around the defined length which corresponds to the total gene size of 1.4 kb + 15 kb = 16.4 kb (see Fig. 3.1.1). Similar observations were made with other pairs of genes, e.g. *Drosophila melanogaster Antennapedia* (Kuroiwa et al., 1985) (see also Table 3.1.1). It may be concluded, therefore, that the correlation (Fig. 3.1.1) previously observed in large numbers of pairs of almost equally sized genes is also seen in pairs of genes different to each other in size. All of the pairs shown in Table ~~#~~ 3.1.1 possess the defined lengths of intergenic distances. Total gene sizes range in length from 1.3 to >100.4 kb (Table 3.1.1). This finding allows us to examine the nature of gene territories in greater detail.

One approach for this examination was provided by a model analysis of hypothetical gene clusters. This analysis was particularly useful for exploring various

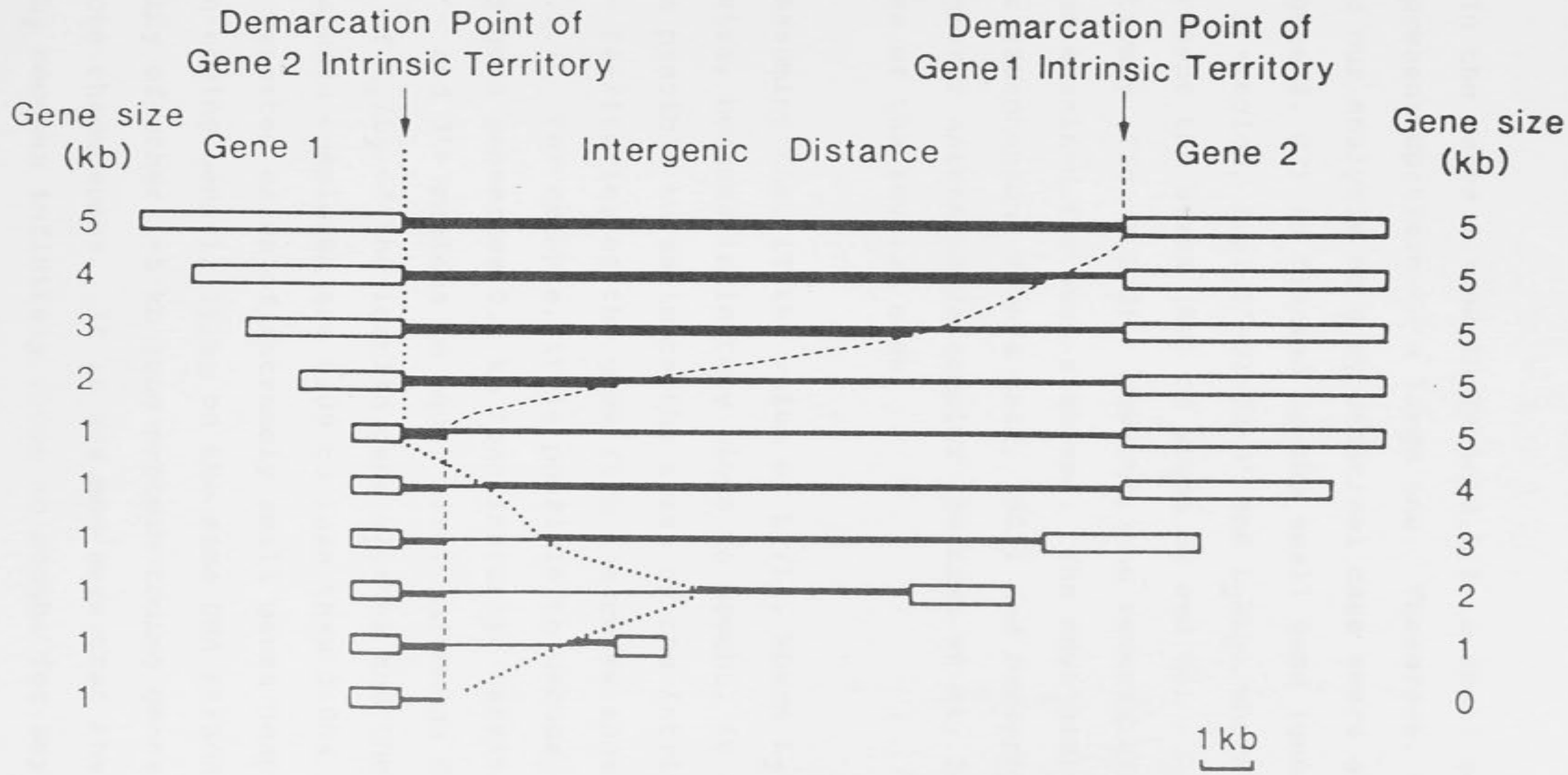
Table 3.1.1
Some examples of pairs of unusual sized genes

gene clusters and gene sizes (kb)	genes, intergenic distances	ratio of genes	total gene sizes (kb)
Newt histone (Stephenson et al., 1981)	←H3 - ←H1 1.11 0.54 0.76	1.41	1.30
Drosophila melanogaster histone (Kedes, 1979)	←H1 - ←H3 1.35 0.95 0.52	1.83	1.47
Xenopus laevis globin (Patient et al., 1980)	→α1 - →β1 7.7 1.2 1.8	1.50	3.0
Human apolipoprotein (Karathanasis, 1985)	→Apo AI - ←Apo CIII - →Apo AIV 12 1.7 1.2	1.41	4.1
Human insulin (Bell et al., 1985)	→Insulin - →Insulin like growth factor II 12.6 1.4 15	10.7	16.4
Drosophila melanogaster bithorax complex (Beachy et al., 1986)	←*Ubx* - ←*bxd* 10 ~75 20	~3.8	~95
Drosophila melanogaster *Anntenapedia* complex (Kuroiwa et al., 1985)	←X - ←*Antp* ~16 ~0.4 >100	>250	>100.4

Arrows show the directions of genes. The numbers show below the genes and intergenic regions of pairs of genes present on the same DNA strand represent the lengths of respective genes and intergenic distances. The relative gene sizes are compared by the ratios of (large gene)/(small gene) within a cluster.

Fig. 3.1.2 A model analysis of various gene clusters.

Pairs of genes, Gene 1 and Gene 2, are present on the same DNA strand and possess the defined lengths of an intergenic distance between them. The defined lengths were obtained from Fig 3.1.1. The gene sizes of Gene 1 vary from 5 kb to 1 kb in length as illustrated on the left side of the figure. Also, the size ~~variance~~ variation of Gene 2 is from 5 kb to 0 kb (see the right side of the figure). The sizes of the intrinsic territories of Gene 1 and Gene 2 were obtained as described in the text and illustrated by bold lines. Demarcation points of the intrinsic territories of Gene 1 and Gene 2 are connected by dashed and dotted lines, respectively. The lengths are according to the scale at the bottom.


Demarcation Point of
Gene 2 Intrinsic Territory
Demarcation Point of
Gene1 Intrinsic Territory
Gene size
(kb)
Gene 1
Intergenic Distance
Gene 2
Gene size
(kb)
5
4
3
2
1
1
1
1
1
5
5
5
5
5
4
3
2
1
0
1 kb

dynamic aspects of a gene territory. To achieve an effective analysis, the following assumptions were made:

(i) In the cases listed in Table 3.1.1, a small gene is often present upstream to a large one. Therefore, we have carried our analysis to a hypothetical case where a large gene (gene 2, G2) is flanked by the small gene (gene 1, G1) at its 5′-region, i.e. 5′-G1-G2-3′ and $L_2 \geq L_1$, where L_2 and L_1 represent the sizes (kb) of genes G2 and G1, respectively (Fig. 3.1.2). However, the reverse situation was also examined for some analyses. The newt histone gene cluster (Stephenson, Erba & Gall, 1981) and *Drosophila melanogaster Antennapedia* complex (Kuroiwa et al, 1985) are examples of the reverse case.

(ii) Assuming that if the value of L_1/L_2, where L_2 is a given size, becomes infinitely close to nought, it would then be possible to estimate the sizes of the intrinsic 5′- and 3′- territories of the gene from the curve shown in Fig. 3.1.1. For example, it is possible to deduce that a 1 kb long gene possesses 0.8 kb long intrinsic territories at both 5′- and 3′- regions in equal size. Although the values of L_1/L_2 of the insulin gene cluster and the *Antennapedia* complexed are 0.09 to less than 0.004, there are no reported cases of extremely small genes (especially protein-coding genes), lying on the same DNA strand within proximity of other 1-5 kb long protein-coding genes on eukaryote chromosomes. It is assumed here that the value of L_1/L_2 becomes infinitely close to nought for any given value of L_2.

dynamic aspects of a gene territory. To achieve an effective analysis, the following assumptions were made:

(i) In the cases listed in Table 3.1.1, a small gene is often present upstream to a large one. Therefore, we have carried our analysis to a hypothetical case where a large gene (gene 2, G2) is flanked by the small gene (gene 1, G1) at its 5′-region, i.e. 5′-G1-G2-3′ and $L_2 \geq L_1$, where L_2 and L_1 represent the sizes (kb) of genes G2 and G1, respectively (Fig. 3.1.2). However, the reverse situation was also examined for some analyses. The newt histone gene cluster (Stephenson, Erba & Gall, 1981) and *Drosophila melanogaster Antennapedia* complex (Kuroiwa et al, 1985) are examples of the reverse case.

(ii) Assuming that if the value of L_1/L_2, where L_2 is a given size, becomes infinitely close to nought, it would then be possible to estimate the sizes of the intrinsic 5′- and 3′- territories of the gene from the curve shown in Fig. 3.1.1. For example, it is possible to deduce that a 1 kb long gene possesses 0.8 kb long intrinsic territories at both 5′- and 3′- regions in equal size. Although the values of L_1/L_2 of the insulin gene cluster and the *Antennapedia* complexed are 0.09 to less than 0.004, there are no reported cases of extremely small genes (especially protein-coding genes), lying on the same DNA strand within proximity of other 1-5 kb long protein-coding genes on eukaryote chromosomes. It is assumed here that the value of L_1/L_2 becomes infinitely close to nought for any given value of L_2.

Selected gene clusters, composed of genes G1 and G2 are given in Fig. 3.1.~~1~~2. In the lower half of the figure L_1 = 1 kb and L_2 varies from 0 to 5 kb, and in the upper half L_2 = 5 kb and L_1 varies from 1 to 5 kb. This figure provides several insights into aspects of the gene territories required for active expression.

An intrinsic territory may be considered as a specific feature of a given gene and is thus likely to be demarcated by a specific conformation, e.g. altered secondary structures. I have examined the possible existence of such secondary structures as "landmarks" using the *Drosophila melanogaster* alcohol dehydrogenase *Adh* gene, which possesses dual (adult and larval) initiation sites (see Chapter 3.2). The adult and larval genes are 1.8 and 1.1 kb long respectively (Benyajati et al., 1983). When the *Adh* gene was under torsional stress, it exhibited at least four S1 nuclease-sensitive sites; at 2.7 kb and 1.4 kb upstream of the adult and larval initiation sites respectively, and 0.8 kb and 3.0. kb downstream from the poly(A) addition site. As some of these sites were also sensitive to T4 endonuclease VII, which specifically attacks Holliday structures (Mizuuchi et al., 1982), this suggests that the *Adh* gene possesses specific conformational features at these sites. The adult and larval *Adh* genes are estimated to possess 2.9 kb and 0.9 kb long intrinsic territories, corresponding to the "total" gene sizes of 1.8 kb (=1.8 kb + 0 kb) and 1.1 kb (=1.1 kb + 0 kb) in Fig. 3.1.1, respectively, at their 5′- and 3′- regions. What is strikingly apparent is that these S1

Table 3.1.2
Sizes of Intrinsic Territories

Genes and sizes (kb)		Theoretical estimation of 5′- and 3′- intrinsic territories (kb)	Experimental estimation of sites sensitive to single strand-specific nucleases* (Distance in kb from termini of the gene). 5′	3′	Ref.
Drosophila melanogaster					
Heat shock protein					
hsp 70 (proximal)	2.3	5.4	~5.0	~5.8	Udvardy *et al.*(1985)
hsp 28	1.1	0.9	0.8**	0.3	Selleck *et al.* (1984)
hsp 23	0.8	0.5	0.2	0.2 or 0.5**	"
Alcohol dehydrogenase					
Adult	1.8	2.9	2.7	3.0	Miyahara *et al.* (unpublished
Larval	1.1	0.9	1.4	0.8	" observation)
Chicken globin					
ρ	1.2	1.1	1.4	none***	Larsen *et al.* (1982)
β^H	~1.5	~1.9	none***	2.2	"
β^A	1.5	1.9	none***	1.3	"
ε	1.6	2.2	1.7	N.D.	"
Sea urchin histone					
H1	~0.6	~0.4	0.6	~0.4	Hentschel (1982)
Human snRNA					
U1	~0.17	0.3	0.3	none***	Htun et al. (1984)

Notes

*: S1 nuclease (in some cases, Bal 31 or *Neurospora crassa* nucleases) - sensitive sites were examined in supercoiled plasmids except for the hsp 70 gene.

**: The site is shared by the paired genes.

***: No sites have been reported which correspond to those predicted. This does not necessarily imply the absolute absence at these regions.

N.D.: Not determined.

nuclease-sensitive sites exist just around or near the presumptive borders of these intrinsic *Adh* gene territories. Similar observations were made with several other genes, including chicken β and β-like globin genes (Larsen & Weintraub, 1982) and *Drosophila melanogaster* heat shock protein genes (Selleck et al., 1984). (Table 3.1.2) Probably, these observations are not just mere coincidence, and the specific conformations I have found may represent the demarcation sites of intrinsic territories.

The existence of landmarks specific to individual genes suggests that the extragenic DNA sequences (e.g. the intrinsic gene territories exhibiting such landmarks) may have evolved together with a given gene (e.g. as in some cases the gene changes in size over long evolutionary periods). Alternatively, it is equally conceivable that many potential landmarks may exist; scattered around at various places on the chromosomes. Genes that have newly evolved (e.g. by gene duplication) or translocated and even genes introduced by transfection experiments may utilize pre-existing landmarks at their new locations for active expression.

In the gene clusters shown in Fig. 3.1.2, the intrinsic territorial regions (estimated from Fig. 3.1.1) are indicated by bold lines. It is quite evident from this figure that the actual intergenic distances between two genes are neither equal to the simple sum of intrinsic territories of paired genes nor simply proportional to gene

size. For example, in a pair of 1 kb long genes the intergenic distance was 150% greater than the simple sum of the two intrinsic territories. The role of this additional sequence in the case of the small gene cluster is not known. This extra region became smaller as one of the paired genes increased in size. When a 1 kb long gene was paired with a 4.5 kb long gene, the gene pair possessed an intergenic distance which is equivalent to the sum of the sizes of the two intrinsic territories assignable to each gene. When one of the paired genes became much larger, the intrinsic territories of both genes partially overlapped. Eventually the intrinsic territories of both genes became fully overlapping when two very large genes were paired, for example, serum albumin α- fetoprotein genes (Kioussis et al., 1981; Ingram, Scott & Tilghman, 1981) and *Ubx* - *bxd* cluster (Beachy, Helfand & Hogness, 1985). Since these large genes are not under territorial confrontation, overlapping at this extent of two intrinsic territories does not obviously interfere with the functions of these large genes.

It seems obvious from the analytical model described above that although the intrinsic territories of a gene are marked by the presence of specific secondary structures, the actual territories of a given gene that are required for active expression are not fixed, but determined in a complex manner as a result of multiple interactions with the territories of neighbouring genes. An intrinsic territory would be eliminated by close occupation of neighbouring genes, if the neighbouring genes come under

territorial confrontation. Consequently, the intergenic sequences between the two genes would be partially deleted so that the intergenic distance would be difined. The observation implies that the potential for gene expression is a function of relations which can be assigned on chromosomes. A complete understanding of the mechanisms underlying territorial confrontation has not yet been achieved.

3.2 Dynamic topology of the *D. melanogaster Adh* gene under torsional stress

3.2.1 Introduction

It has often been reported that sites hypersensitive to S1 nuclease exist around promoter regions or upstream from 5′- ends of genes in recombinant DNA (Iatron & Tsitilon, 1983; Goding & Russell, 1983; Larson & Weintraub, 1982; Mokern et al., 1984; Cristophe et al., 1985; Htu et al., 1984; Margot & Hardison, 1985; Selleck et al, 1984). The hypersensitive sites can be detected on negatively supercoiled DNA but not on linearised DNA. In a supercoiled DNA molecule, torsional stress leads certain DNA sequences to take up non-B form structures or altered secondary structures that are susceptible to digestion with S1 nuclease. A duplex DNA molecule does not possess an ideal B-form structure throughout its whole molecule but its configuration is heterogeneous. It also seems that a duplex DNA molecule (supercoiled) can elastically alter its configuration according to the strength of its torsional stress.

The superhelicity of a closed-circular DNA molecule is expressed in the following equation:

$$\mathbf{Lk = Tw + Wr} \qquad (1)$$

Here **Tw** is the total molecular twist; the number of times one strand rotates around the duplex central axis. **Wr** is

the writhing number; a measure of the global tertiary structure. **Lk** is the molecular linking number; the number of times one strand links through the closed circle formed by the other strand (Wang et al., 1982).

The parameters Tw and Wr can not be measured experimentally for supercoiled DNA at the present time. In most cases, the superhelical density (σ) is often used to express the superhelicity of supercoiled DNA. The definition of superhelical density is:

$$\sigma = (Lk - Lk^{\circ})/Lk^{\circ} \qquad (2)$$

For a relaxed DNA, it appears that Wr = 0 and Tw = Lk°. Since the helical periodicity of relaxed DNA is known to be approximately 10.5 bp per one turn, the superhelical density means the number of superhelical turns per 10.5 bp.

Since the DNA molecule is covalently closed into a circle, Lk is fixed, while the values of Wr and Tw can be varied, i.e. superhelical constraint, by assuring the constancy of Lk, topologically couples the secondary and tertiary structures of the molecule. Changes in Wr involve bending, whereas changes in Tw may include torsional deformations of existing secondary structures and stress-induced local transitions to alternative conformations. Such conformational changes have been considered theoretically as a function of torsional energy (Vologodskii et al., 1979; Benham, 1979, 1981, 1982). Experimental studies on electrophoretic mobility and

nuclease sensitivity demonstrated that the free energy of negative supercoiling caused a right-to-left handed helical transition at a defined level of torsional stress (Singleton et al., 1982; Peck et al., 1982). Cruciform structure is also induced energetically at a specific superhelical density (Singleton & Wells, 1982a; Courey & Wang, 1983).

In this work, an attempt was made to characterise the structures around the *D. melanogaster Adh* gene region induced by torsional stress under various conditions. The aim of such experiments was to investigate the structural basis for gene territoriality in relation to the plasticity of DNA molecules. First, some specific sites hypersensitive to S1 nuclease were found on a supercoiled recombinant DNA containing the *Adh* gene. These sites were induced by respective degrees of torsional stress. Secondly, I analysed the susceptibility of the DNA to various other enzymes; single strand specific nucleases (mung bean and *N. crassa* nuclease), Bal 31 nuclease, phosphodiesterase I and T4 endonuclease VII. Single strand binding protein (SSB) purified from *E. coli* was also used to probe single stranded regions in the *Adh* DNA (see Table 3.2.1, for detail). The results here suggest that the DNA molecule around the *Adh* gene region is organised into discrete functional sequences possessing characteristic structural properties. They imply that the borders of the intrinsic territory of the *Adh* gene may be marked by sequences having unusual conformational properties.

3.2.2 Materials and methods

(a) Enzymes and reagents

S1 nuclease from *Aspergillus oryzae* was purchased from Sigma. Mung bean nuclease, *N. crassa* nuclease and single strand binding protein from *E. coli* were purchased from P.L. Biochemicals. T4 endonuclease VII was a gift of D. Kemper. Topoisomerase I from wheat germ and Bal 31 nuclease were purchased from BRL.

(b) Digestion of supercoiled plasmid DNA

All nuclease digestions were performed at 30°C for 20 min. S1 nuclease digestion was carried out in a buffer of 30 mM Na-acetate (pH 4.5), 3 mM $ZnSO_4$, 0.2 mM EDTA and 50 mM NaCl. *N. crassa* nuclease digestion was carried out in a buffer of 10 mM Hepes (pH 7.9), 0.33 mM DTT, 7.5 mM Mg-acetate and 60 mM KCl, phosphodiesterase I digestion in 25 mM Tris-Cl (ph 9.2), 0.5 mM Mg-acetate and 0.00625% Triton X-100, Bal 31 nuclease digestion in 20 mM Tris-Cl (pH 8.0), 200 mM NaCl, 12 mM $MgCl_2$ and 12 mM $CaCl_2$ and endonuclease VII digestion in 50 mM Tris-Cl (pH 8.0), 10 mM $MgCl_2$, 10 mM 2-mercaptoethanol and 250 µg/ml bovine serum albumin. These nucleases cut both strands of DNA to produce linearised DNA. However, mung bean nuclease cuts only a single strand of DNA or nicks it in 10 mM Tris-Cl (pH 7.0). To detect the sites susceptible to mung bean nuclease digestion, after a DNA treated with mung bean nuclease was completely linearised with Eco RI then the other strands at nicked

sites were cut by S1 nuclease. For S1 nuclease can recognise nicked sites and cut opposite strands, but does not digest any other region of linearised DNA molecules (see Fig. 3.2.9).

These nuclease treated DNA molecules with double strands cuts were further restricted by an appropriate enzyme, mainly Eco RI, then separated by electrophoresis on 1% agarose gels, followed by Southern hybridization to map the sites susceptible to nucleases. As probes for indirect labeling (Wu, 1980), a 150 bp Pvu I - Eco RI fragment was used to detect the *Adh* gene region (Fig 3.2.1), and a 150 bp Pvu II - Bam HI fragment (Bam HI upstream probe) and a 250 bp fragment (Bam HI downstream probe) were used to detect regions far upstream and downstream from the Bam HI site (Fig 3.2.14). Furthermore, a 150 bp Eco RI - Hind III fragment (Fig 3.2.1) in the *Adh* gene region was also used as a probe to confirm the previous results. These fragments were labelled with [α-^{32}P] dATP by a random priming procedure (see Chapter 2).

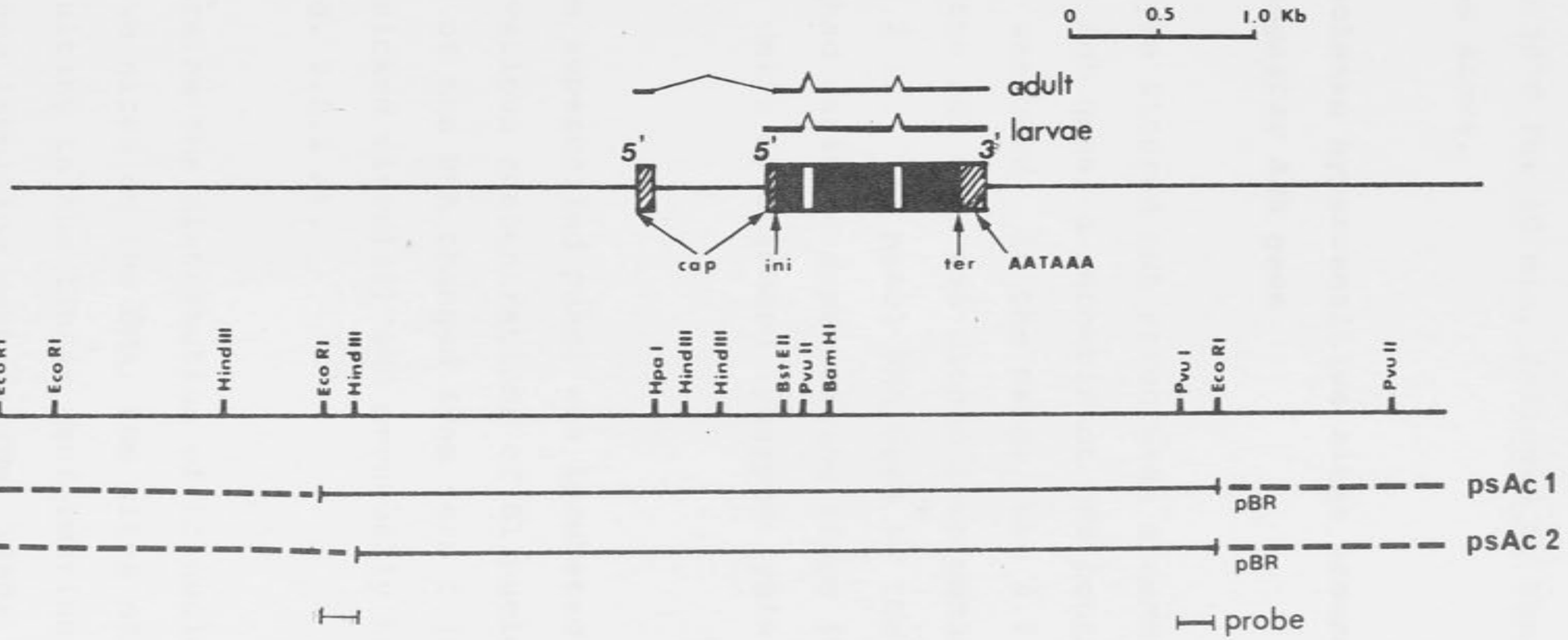
(c) Effects of single strand binding protein and spermidine on S1 nuclease digestion

A physiologically supercoiled recombinant DNA, psAc2, was first incubated in 10 µl of 10 mM Hepes (pH 7.9), 0.33 mM DTT, 7.5 mM Mg-acetate, 60 mM KCl and a varied concentration of single strand binding protein or spermidine at 30°C for 20 min, then 10 x S1 nuclease buffer was added to adjust to 1 x buffer and S1 nuclease was added

Fig. 3.2.1 Schematic map of the *Drosophila melanogaster Adh* gene locus.

Filled and hatched boxes represent translated and non-translated regions in exons. Two different transcripts (adult and larval) are indicated by lines above the gene. The DNA probes used for indirect end-labeling are indicated below the gene. In psAc1 and psAc2 plasmid DNA, 4.7 kb EcoRI and 4.55 kb HindIII-EcoRI fragments were cloned into pBR322 vector, respectively.

D. melanogaster Adh gene



at a varied concentration. This mixture was further incubated at 30°C for 10 min, followed by the same procedures as above.

3.2.3 S1 nuclease hypersensitive sites around the *D. melanogaster Adh* gene

To analyse altered DNA structures around the *D. melanogaster Adh* gene, a recombinant DNA containing the *Adh* gene, psAc1, was used. In the psAc1, the 4.7 kb Eco RI fragment of the *Adh* gene was cloned into pBR322 (Goldberg, 1980; Fig. 3.2.1). The psAc1 DNA used in the present experiments had negative supercoiling since it was prepared from *E. coli* cells by standard lysozyme lysis methods (see Chapter 2).

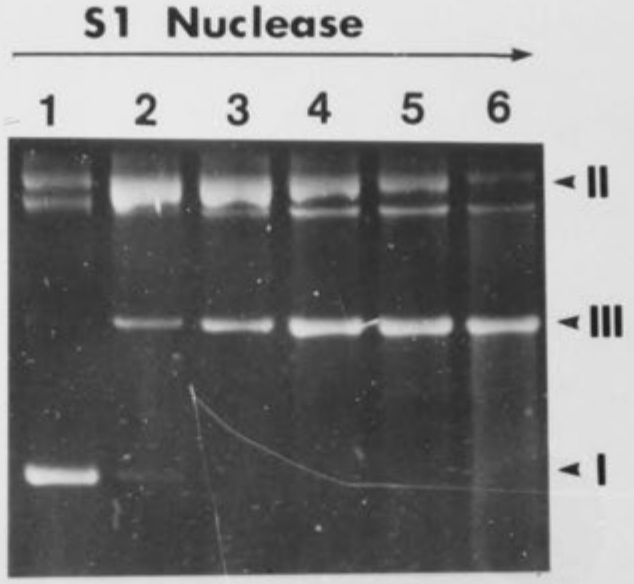
When the supercoiled psAc1 was incubated in the presence of various concentrations of S1 nuclease, the conformation of the DNA changed from form I (supercoiled) to form II (nicked circular) and eventually to form III (linear) (Fig. 3.2.2 A).

To determine the distribution of S1 nuclease hypersensitive sites on the DNA, the sites of double strand cleavage resulting in the linear conformation were mapped by indirect end labelling methods (Wu, 1980). The S1 nuclease-treated DNAs were further cleaved by Eco RI and separated by electrophoresis on a 1% agarose gel. They were Southern-transferred and hybridized with a ^{32}P-labelled Eco RI - Pvu I fragment (~150 bp) (Fig 3.2.1).

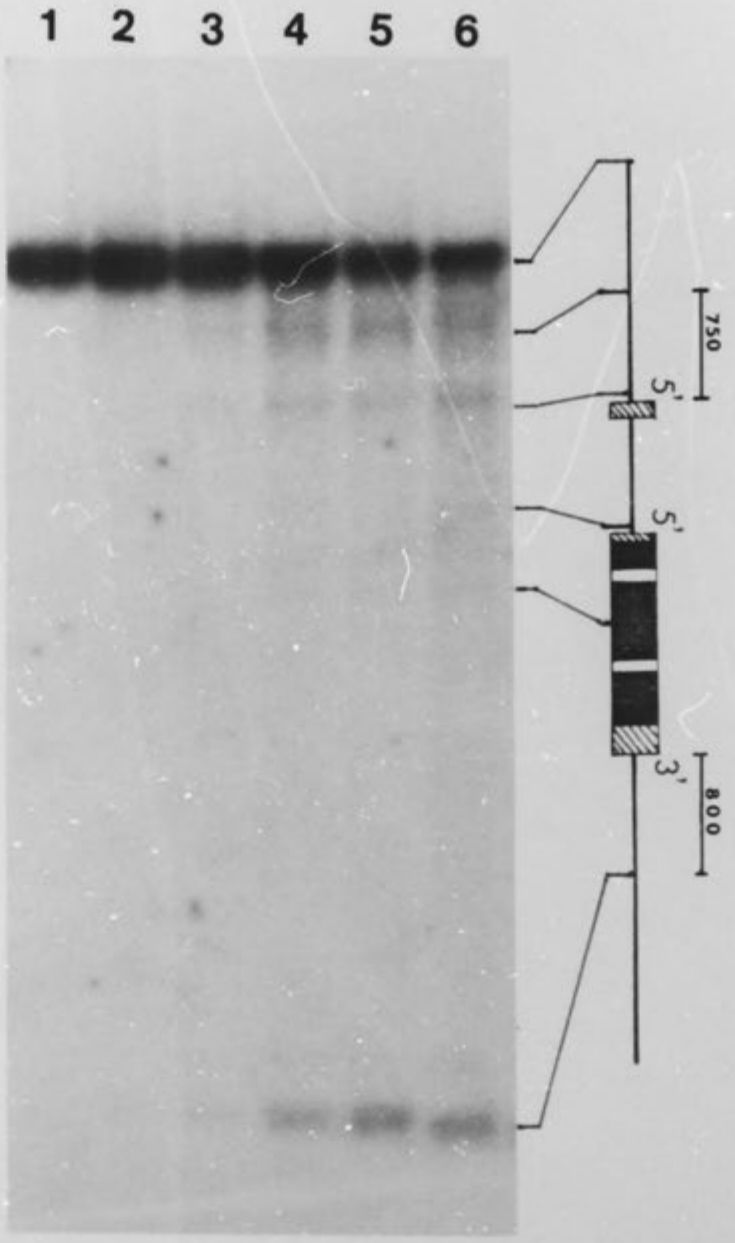
Fig. 3.2.2 Distribution of S1 nuclease hypersensitive sites at the *Adh* locus on the supercoiled DNA.

Supercoiled *Adh* DNA, psAc1, was digested at 30°C for 20 min with S1 nuclease at various concentrations; 0 (lane 1), 50 (lane 2), 125 (lane 3), 250 (lane 4), 500 (lane 5), 1,000 (lane 6) units/ml, respectively. (A) The S1-digested DNA was electrophoresed on 1% agarose gel. The ethidium bromide-stained gel is shown. I; supercoiled, II; nicked circular, III; linear. (B) The S1 digested DNA was further cleaved with EcoRI and electrophoresed on 1% agarose gel, followed by indirect end-labeling methods. The PvuI-EcoRI fragment was used as a probe. The prominent bands are indicated by lines connecting to the corresponding sites on the map of the *Adh* gene.

A



B



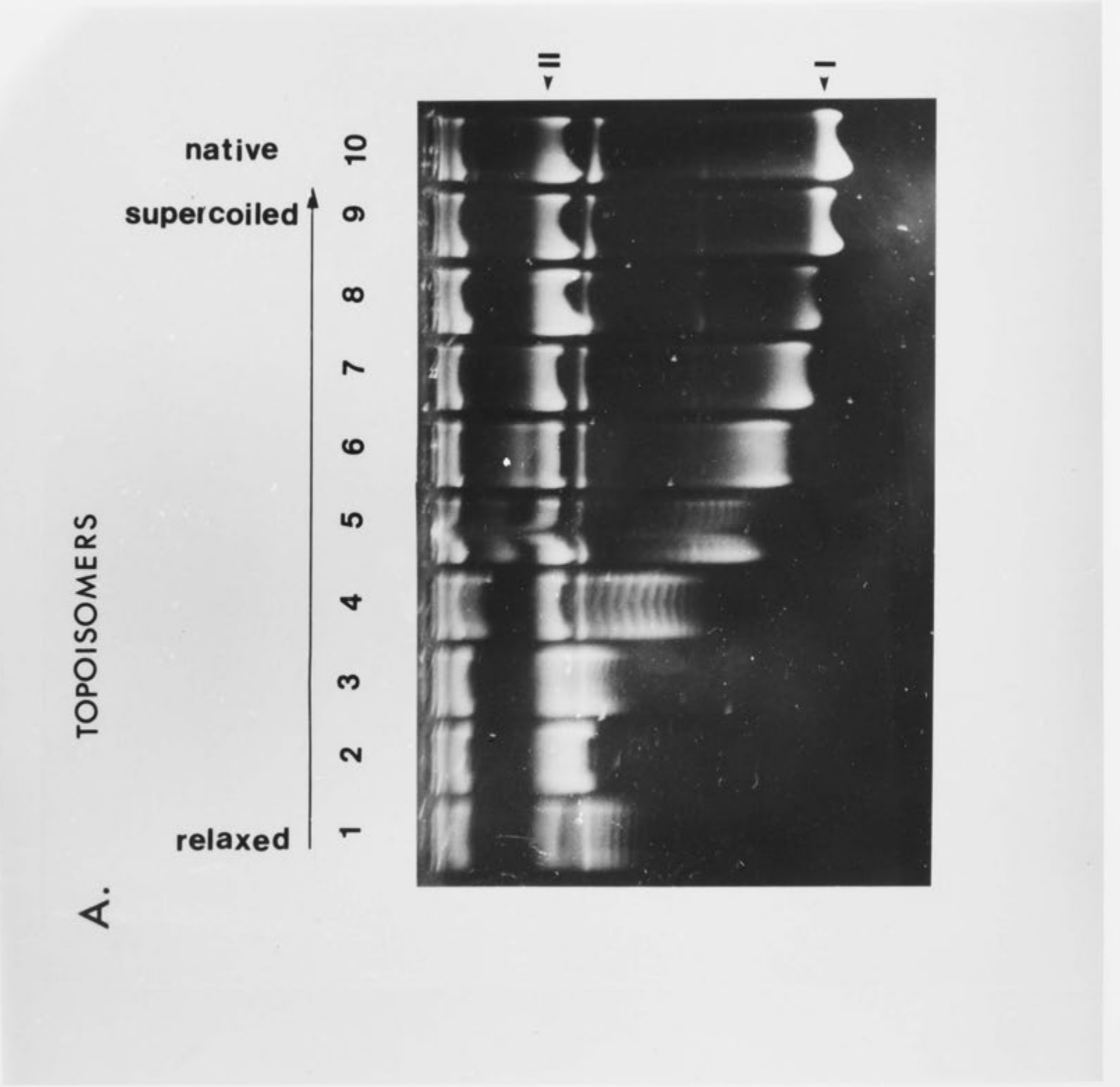
The location of each cleavage site was given directly by the length relative to the Eco RI site and its sensitivity was also given by the intensity of detected band on an autoradiograph.

On the supercoiled DNA, three major sites hypersensitive to S1 nuclease were detectable. One is located 750 bp upstream from the adult capping site, the second one is located around the adult promoter site and the third one is located 800 bp downstream from the 3′- end (Fig. 3.2.2 B). Several other sites were also detectable but less sensitive to S1 nuclease; around the larval promoter, within the first intron, 690 bp downstream from the 3′- end and around the poly(A) addition site (Fig. 3.2.2 B). Almost the same pattern of S1 nuclease cleavage was obtained reproducibly using different batches of psAc1 DNA, and the location of these hypersensitive sites was confirmed using different probes and different restriction enzyme cutting (data not shown). These sites were not detected at all when the psAc1 DNA were linearised by Eco RI before digestion with S1 nuclease. Therefore, it seems that the DNA sequences at these sites were actually induced to be susceptible to S1 cleavage by DNA supercoiling or torsional stress. Some sites correspond to functional sites of the *Adh* gene. This result suggested that various functional sites of a gene have characteristic structural properties.

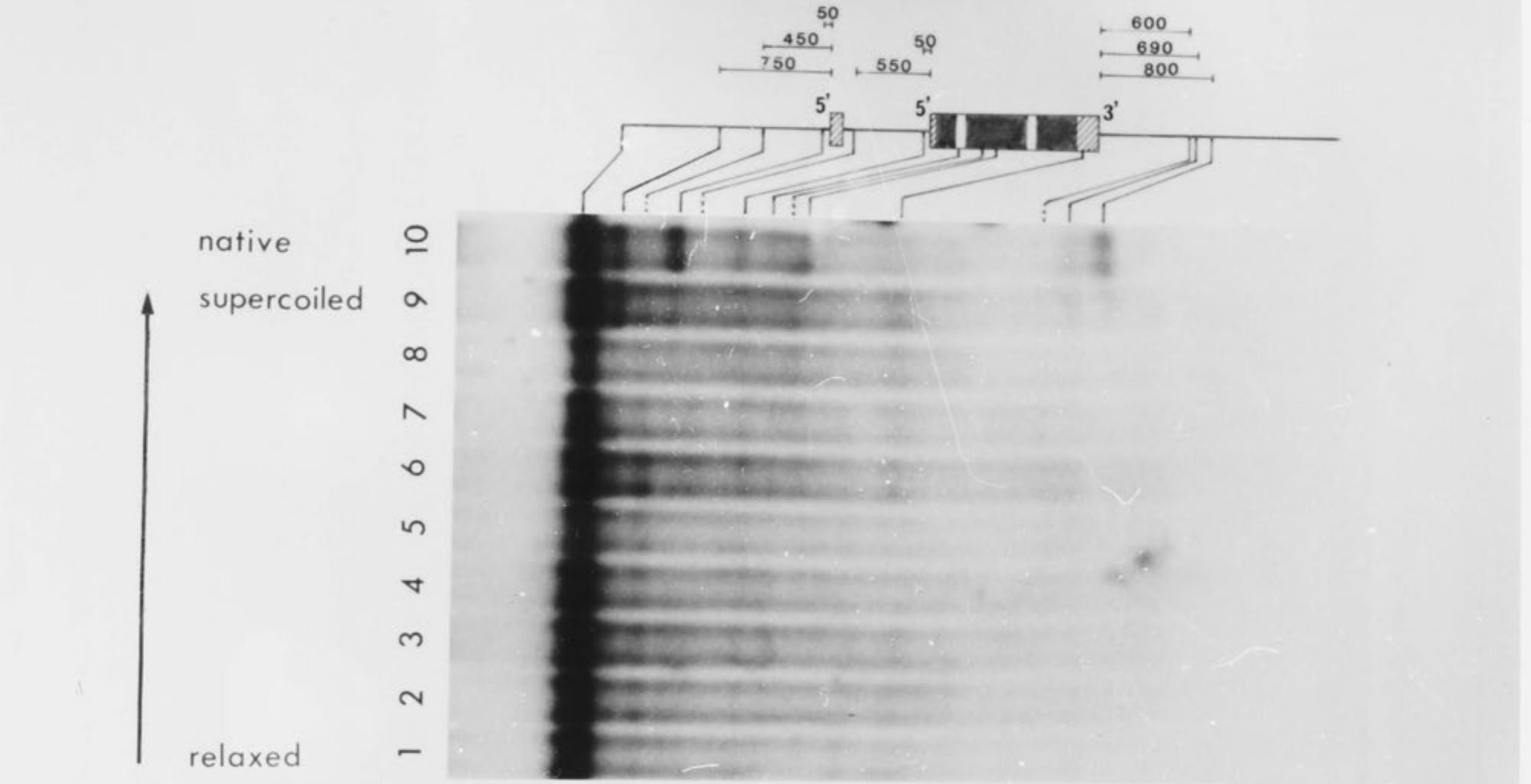
3.2.4 Elastic alteration of DNA conformation according to torsional stress

Fig. 3.2.3 Dependence of distribution of S1 nuclease hypersensitive sites on superhelical densities of topoisomers.

The topoisomers were prepared by treatment of psAc1 DNA with topoisomerase I at 30°C in the presense of various concentrations of ethidium bromide; 0, 1, 2, 3, 4, 5, 6, 7 and 8 μM. (A) The psAc1 topoisomers and physiologically supercoiled psAc1 were purified and separated by electrophoresis on 1% agarose gel. The average superhelical densities ($-\sigma$) of those topoisomers were estimated to be 0 (lane 1), 0.007 (lane 2), 0.013 (lane 3), 0.019 (lane 4), 0.025 (lane 5), 0.031 (lane 6), 0.037 (lane 7), 0.043 (lane 8), 0.049 (lane 9) and 0.055 (physiological; lane 10), respectively. (B) The psAc1 topoisomers were digested at 30°C for 20 min with S1 nuclease at 1,000 units/ml. The S1-digested topoisomers were further cleaved by EcoRI and electrophoresed on 1% agarose gel followed by indirect end-labeling methods. The PvuI-EcoRI fragment was used as a probe. The S1 hypersensitive sites were mapped and indicated by lines connecting to the schematic map of the *Adh* gene.


A.
TOPOISOMERS
relaxed
native
supercoiled
1
2
3
4
5
6
7
8
9
10
II
I

B.


50
450
750
50
550
600
690
800
5'
5'
3'
native
supercoiled
relaxed
1 2 3 4 5 6 7 8 9 10

The formation of a given cruciform or Z-form structure has been shown to occur transiently at a specific superhelical density (Singleton & Wells, 1982a; Singleton et al., 1982). It is likely that such a conformational transition depends on other transitions which occur simultaneously or consecutively within the circular DNA molecule. In such a case, these conformational transitions should compete for the free energy of supercoiling. Consequently, altered secondary DNA structures in specific regions would form in a hierarchical manner.

This raises a question as to whether the S1 nuclease hypersensitive sites observed around the *Adh* gene in the psAc1 DNA alter according to the degree of torsional stress to which the entire DNA molecule is subjected. I aimed here to clarify this question. Topoisomers of psAc1 were prepared by treatment with ethidium bromide and topoisomerase I as described in Chapter 2. The average superhelical density ($-\sigma$) in the topoisomers ranged from 0 (relaxed) to 0.055 (physiological) (Fig. 3.2.3 A). The topoisomers with lower superhelical density were more resistant to digestion by S1 nuclease; most topoisomers still remained even after the digestion, and a small fraction of them was converted to linear or nicked forms (data not shown). Sites hypersensitive to S1 nuclease on the psAc1 topoisomers were detected as in previous experiments (Fig. 3.2.3 B). Of particular interest was that the S1 nuclease sensitivity at specific sites showed a superhelical density-dependence. Firstly, the three major hypersensitive sites present on the physiologically

supercoiled DNA rapidly diminished in sensitivity as the superhelical density of the DNA became lower. They were undetectable on the psAc1 topoisomers with superhelical density ($-\sigma$) lower than 0.04. On the other hand, some other hypersensitive sites on the physiologically supercoiled DNA were detectable even on relaxed DNA, i.e. around the larval promoter, around the poly(A) addition site, and 690 bp from the 3'- end. Secondly, it should be noted here that at least four new S1 nuclease hypersensitive sites, which were undetectable on the physiologically supercoiled DNA, became detectable when the superhelical density dropped below the physiological value. Two sites, mapped 800 bp and 450 bp upstream from the distal cap site, remained detectable but very faint on the relaxed DNA. However, two other sites, mapped at the centre of the *Adh* coding region and 600 bp downstream from the 3'- end , were detected only on the moderately supercoiled DNA, and not on either the relaxed or highly supercoiled DNA. I also prepared psAc1-topoisomers with higher superhelical densities and analysed S1 hypersensitive sites in the same way as above. The results showed that the patterns of the S1 hypersensitive sites on three such topoisomers were exactly the same as that of physiologically supercoiled DNA (data not shown).

Thus it was concluded that the hypersensitivities to S1 nuclease at various sites elastically altered in a complex manner depending on the superhelical density of the circular DNA molecules. That is, each topoisomer of psAc1 DNA possesses specific altered secondary structures.

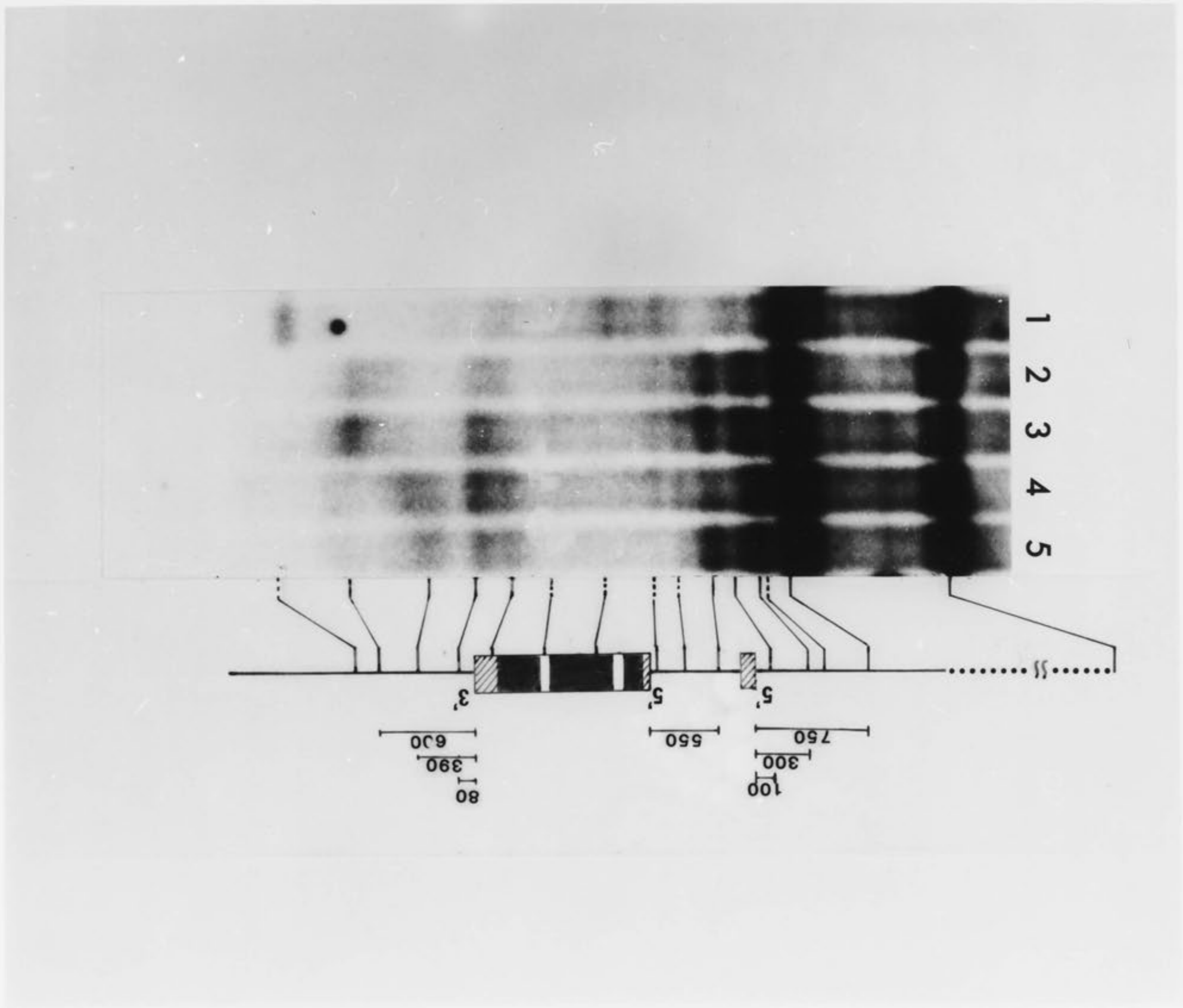
3.2.5 Recognition of altered DNA structures under torsional stress by single strand binding protein

The *E. coli* single strand binding protein (SSB) can stabilise denaturation 'bubbles' by binding to single-stranded regions of DNA (Sigal et al., 1972; Kowalczykowski et al., 1981). Consequently, the stabilization of DNA bubbles by the SSB releases the torsional stress of a supercoiled circular DNA molecule, but it does not result in any change in the linking number of the DNA.

To examine the conformational alteration of DNA around the *Adh* gene region further, I analysed the effects of the SSB on the pattern of S1 nuclease hypersensitive sites on a recombinant DNA, psAc2, which contains a 4.6 kb Hind III - Eco RI fragment of the *Adh* gene region in pBR322 (Fig. 3.2.1). The physiologically supercoiled psAc2 DNA was first allowed to bind to varied amounts of SSB (10 µg/ml to 200 µg/ml), followed by digestion with S1 nuclease, and then the sites susceptible to the digestion were determined by indirect end labelling methods as mentioned in Chapter 3.2.2. The results are shown in Fig 3.2.4. Apparently, the patterns of S1 hypersensitive sites obtained around the *Adh* gene region were changed by the addition of SSB. A few distinct sensitive sites, which were undetected in the absence of SSB, appeared on the addition of a small amount of SSB in both the 5'- and 3'- flanking regions of the *Adh* gene (Fig 3.2.4, lines 2 and 3). It was noticed that these new sites seemed to be the same as the hypersensitive sites which were detected only on the slightly relaxed DNA, i.e.

Fig. 3.2.4 Effects of single strand binding protein on the distribution of S1 nuclease hypersensitive sites.

The supercoiled psAc2 DNA was first incubated at 30°C for 20 min with SSB at various concentrations of 0 (lane 1), 25 (lane 2), 50 (lane 3), 100 (lane 4) and 200 µg/ml (lane 5), respectively, then digested at 30°C for 10 min with 500 units/ml S1 nuclease. The treated DNA was further digested with EcoRI and electrophoresed on 1% agarose gel, followed by indirect end-labeling method. The PvuI-EcoRI fragment was used as a probe.


1
2
3
4
5
5'
3'
5'
3'
100
300
750
550
80
390
630

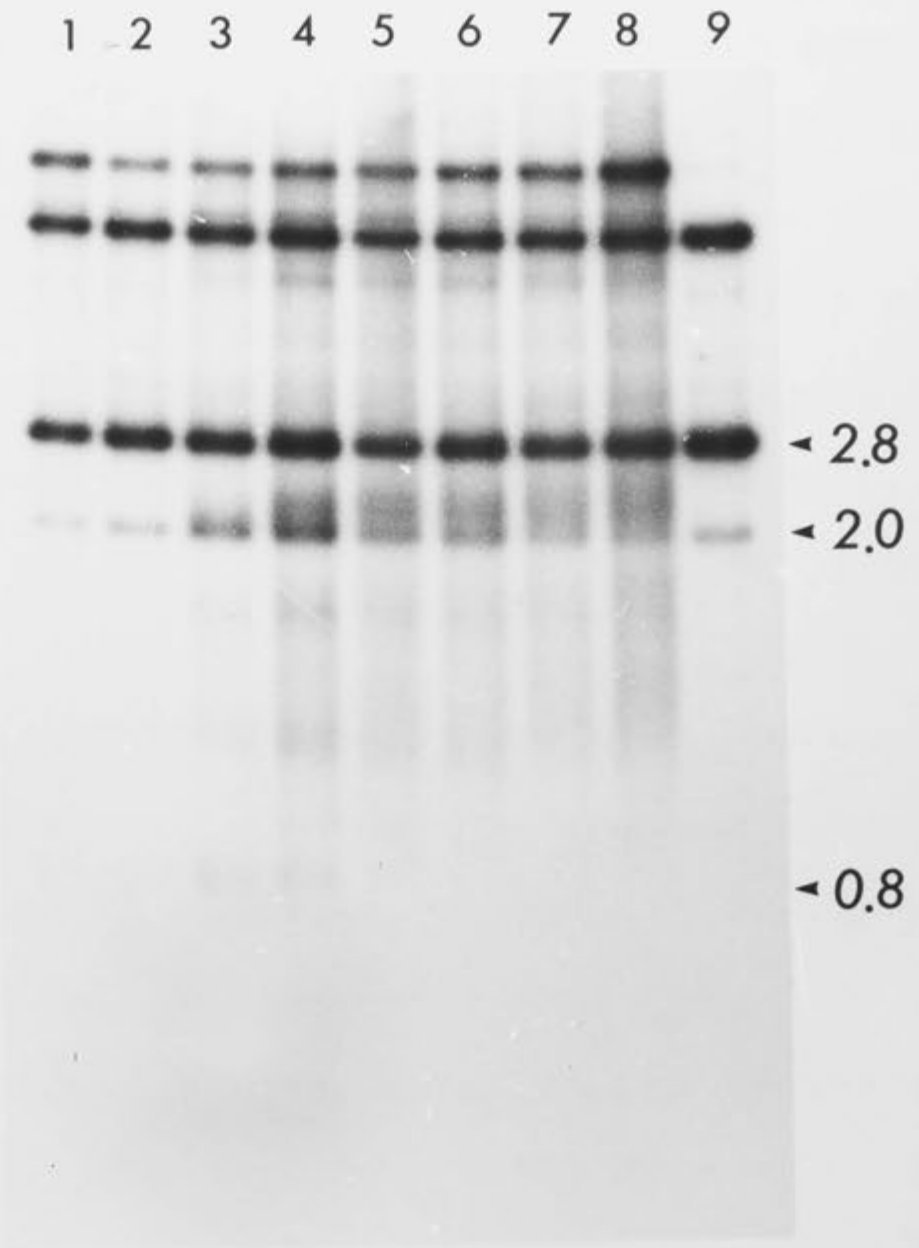
550 bp upstream from the larval capping site and 600 bp downstream from the 3′- end (see also Fig. 3.2.3). ~~concomitance~~ Concomitantly with the appearance of the new sites, however, some of the hypersensitive sites which were detected in the absence of SSB became undetectable upon the addition of SSB. They were mapped in the second exon and 800 bp downstream from the 3′- end of the *Adh* gene. Upon the addition of higher amounts of SSB, the pattern of S1 hypersensitive sites changed further in a complex manner. The hypersensitive sites newly detected at a low level of SSB did become more susceptible to digestion with S1 nuclease upon the addition of more SSB, but some of them, unexpectantly, became undetectable. In addition, some novel sites appeared, previously undetectable on any topoisomers of the *Adh* gene in the absence of SSB. It is possible that specific altered structures are formed at these sites under torsional stress, but are not detectable by S1 nuclease alone. However, they are readily recognised by SSB at the higher concentrations. Alternatively, it also is possible that the binding of SSB to one region indirectly affects the conformation of DNA at another region on the circular DNA.

Of all the previously described sites only the site 750 bp upstream from the adult capping site became more sensitive to S1 nuclease digestion as the amount of added SSB increased (Fig. 3.2.4). This particular hypersensitive site seemed to have a unique feature and therefore it was examined in more detail. The supercoiled psAc2 DNA was treated with S1 nuclease in the presence of SSB (100 µg/ml)

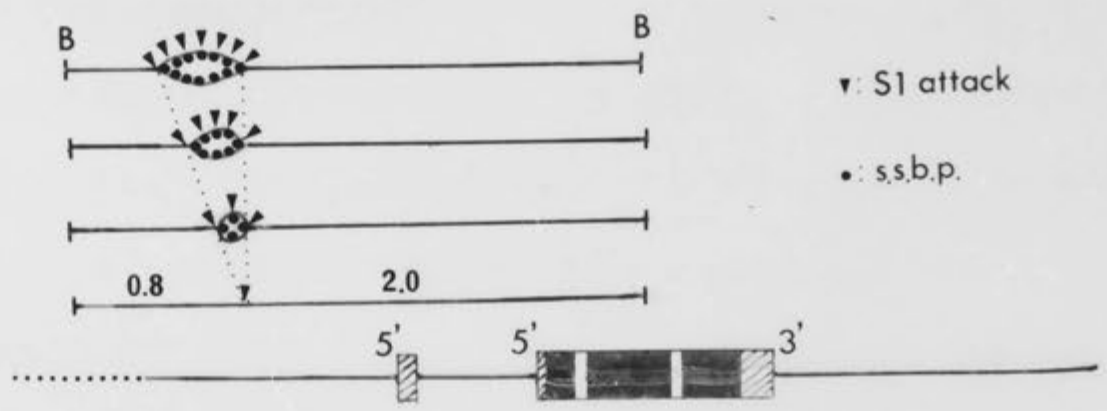
Fig. 3.2.5 Unidirectional binding of single strand binding protein to the upstream region of the *Adh* gene.

(A) The supercoiled *Adh* DNA, psAc2, was incubated at 30°C for 20 min with SSB at various concentrations of 0 (lane 1), 10 (lane 2), 25 (lane 3), 50 (lane 4), 75 (lane 5), 100 (lane 6), 150 (lane 7) and 200 μg/ml (lane 8), respectively, and then digested at 30°C for 15 min with 500 units/ml S1 nuclease. The treated DNA was further cleaved by Bal 31 and the BamHI site was labeled with ^{32}P by Klenow fragment reaction. The labeled fragments were electrophoresed on 1.5% agarose gel, then the gel was dried on a 3MM paper. (B) The model of unidirectional binding of SSB is schematically shown. An altered DNA structure is first recognized by SSB, then the 'denaturation bubble' is stabilized. The bubble is unidirectionally extended upstream by cooperative binding of SSB. ●, a molecule of SSB; ▼, the S1 nuclease attack.

A



B



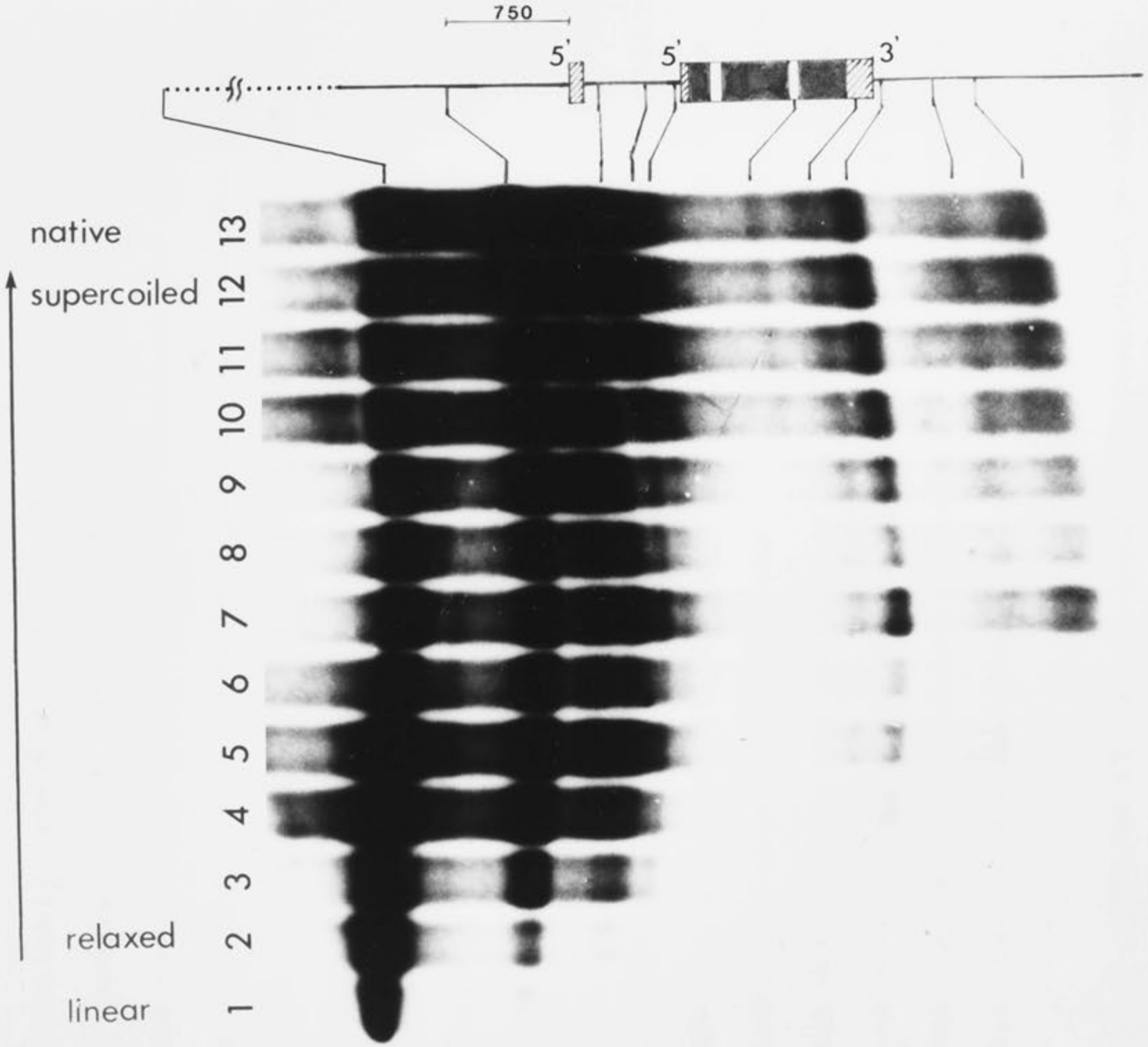
as above, then restricted by Bam HI, the site of which was located near the hypersensitive site. The resulting fragments were labelled at the 3′- end of the Bam HI sites with ^{32}P, then separated on agarose gels. The result was shown in Fig. 3.2.5. The electrophoretic band corresponding to the 2 kb long DNA fragment, the hypersensitive site of which was mapped 750 bp upstream from the distal capping site, became unidirectionally broader in proportion to the increase in the amount of the added SSB (Fig. 3.2.5 A). This observation indicates the unidirectional formation of a DNA-SSB complex at the site. This is illustrated schematically in Fig. 3.2.5 B. Thus, this particular region 5′- upstream from the *Adh* gene possesses a special feature in its DNA sequences, which probably leads to the unidirectional binding of SSB.

Another interesting observation with SSB was that SSB renders the extragenic (non-coding) regions around the *Adh* gene sensitive to S1 nuclease, whereas it leaves the intragenic (coding) region rather resistant. It suggests that the conformational properties of DNA sequences at the intragenic region are distinguishable from those in the extragenic regions.

I also used the psAc2-topoisomers to examine the effects of SSB on the S1 hypersensitive sites around the *Adh* gene (Fig. 3.2.6). The susceptibility of psAc2 DNA to S1 nuclease with SSB became lower as the superhelical density decreased. Of particular interest was that no sites were detectable on the fully relaxed DNA molecule,

Fig. 3.2.6 Supercoil-dependent binding of single strand binding protein to topoisomers.

The topoisomers were prepared by treatment of psAc2 DNA with topoisomerase I at 30°C in the presense of ethidium bromide of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10 μM. The average supercoiled densities (-σ) were determined by separation on chloroquine-agarose gel (Shure et al., 1977); 0, 0.008, 0.016, 0.024, 0.032, 0.040, 0.048, 0.056, 0.064, 0.072 and 0.080. The psAc2-topoisomers (lane 1 to 11) and physiologically supercoiled psAc2 (lane12) were incubated at 30°C for 20 min with 100 μg/ml SSB, and then digested at 30°C for 15 min with 500 units/ml S1 nuclease. The treated DNA was further cleaved by EcoRI, and electrophoresed on 1% agarose gel, followed by indirect end-labeling methods. The PvuI-EcoRI fragment was used as a probe. The S1 hypersensitive sites are indicated by bars connecting to the corresponding sites on the schematic map of the *Adh* gene.


750
5'
5'
3'
native
supercoiled
relaxed
linear
13
12
11
10
9
8
7
6
5
4
3
2
1

even in the presence of SSB. Probably, SSB was unable to bind to the fully relaxed DNA. This observation clearly states that only altered DNA structures induced under torsional stress are recognised by SSB.

In the presence of SSB, the digestion patterns did not change significantly among psAc2-topoisomer DNAs, whereas the pattern of S1 hypersensitive sites did, as shown previously (see Fig. 3.2.2). The recognition of DNA structures by SSB may be different from that by S1 nuclease, albeit both SSB and S1 nuclease are referred to as 'single strand specific', possibly because the denatured state of DNA driven by SSB is metastable. In contrast, S1 nuclease will recognise and cleave the transient denaturation bubbles, which are generated by thermal fluctuations in the supercoiled DNA.

I observed, indeed, that not only nicked and linear DNA but also small fragmented DNA was detected when psAc1 DNA was digested with S1 nuclease in the presence of SSB (data not shown). In contrast, only linear and nicked DNA was detected in the absence of SSB. This result probably indicates that S1 can still recognise the sites of SSB - DNA complex which is metastable even after the supercoiling was lost by a primary nicking.

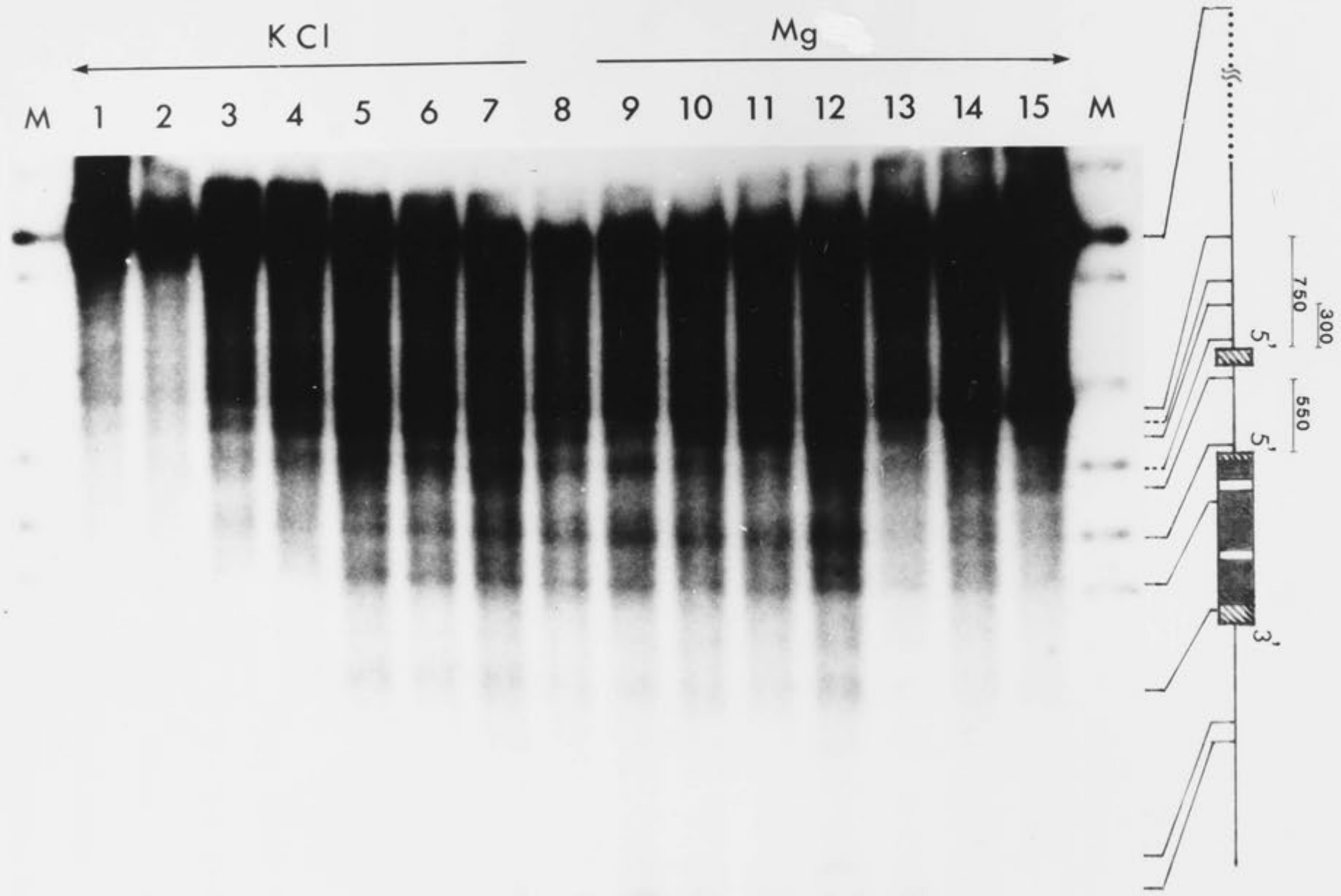
3.2.6 Effects of ion strength and spermidine on altered DNA structures

It is well known that various ions at different

strengths can significantly affect the conformation of DNA molecules in solution (Saenger, 1983). For example, increasing the concentrations of alkali metal ions shifts the induced helix-coil transitions of duplex DNAs to higher temperatures, probably because repulsion forces between negatively charged phosphate groups are neutralized by cations. Metal ions, however, can also have destabilising effect on the DNA double-helical structure if they interact with bases rather than with phosphate groups. Furthermore, in cells, ion concentrations play important roles in cellular functions. Specifically, chromatin structures are well known to change dynamically due to the ion concentrations (McGhee & Felsenfeld, 1980). Therefore, an attempt was made to examine the effects of ions and other materials on the conformational alteration of DNA of the *Adh* gene region. Addition of KCl at various concentration (0 to 200 mM) to the S1 nuclease reaction resulted in a decrease of sensitivity to S1 nuclease on the supercoiled psAc2 DNA. It, however, did not significantly change the patterns of S1 hypersensitive sites on the DNA (Fig. 3.2.7). On the other hand, Mg(II) ion indeed induced an alteration of the pattern of S1 hypersensitive sites on the DNA (Fig. 3.2.7). As described previously (Chapter 3.2.1), the three major S1 hypersensitive sites were detectable without Mg(II) ion. However, two of them became less sensitive to S1 nuclease on addition of Mg(II) ion and eventually undetectable at a concentration of 2 mM; i.e. the adult capping site and 800 bp downstream from the 3'-end of the *Adh* gene. Interestingly, another major site, 750 bp upstream of the adult capping site, became

Fig. 3.2.7 Effects of K^+ or Mg^{2+} ions on the distribution of S1 nuclease hypersensitive sites at the *Adh* locus.

The supercoiled *Adh* DNA, psAc2, was incubated at 30°C for 10 min with various concentrations of KCl; 5 (lane 7), 10 (lane 6), 25 (lane 5), 50 (lane 4), 100 (lane 3), 150 (lane 2) and 200 mM (lane 1), or Mg-acetate; 0.2 (lane 9), 0.5 (lane 10), 1.0 (lane 11), 2.0 (lane 12), 5.0 (lane 13), 10 (lane 14) and 20 mM (lane 15), and without KCl and Mg-acetate (lane 8). Then the DNA was digested at 30°C for 10 min with 125 units/ml S1 nuclease. The S1 hypersensitive sites were detected as described in the legend to Fig. 3.2.4.


KCl
Mg
M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 M
300
750
550
5'
3'

significantly more sensitive on the addition of Mg(II). Thus, the major sites, which were detected in the absence of Mg(II) ion, showed different modes of response to Mg(II) ion effects on their sensitivity to S1 nuclease. This observation implies that different types of altered DNA structures are formed by torsional stress under the standard conditions at the respective major sites.

In addition to the major sites, Mg(II) ion also gave effects on other sites. On the addition of Mg(II) ion, the following sites became sensitive to S1 nuclease; 550 bp downstream from the larval capping site, and 500 bp and 300 bp upstream from the adult capping site (Fig. 3.2.7). It seems likely that the two former sites were identical to those detected on the slightly relaxed DNA by S1 nuclease in the previous experiment (see Fig. 3.2.2). It would indicate that these sites possesses a plastic conformational structure, which easily alters in various microenvironments.

Spermidine is one of the polyamines, which are polycationic compounds naturally present in nuclei. Polyamines can bind to DNA electrostatically and change the conformation of DNA (Gasule & Schellman, 1976; Drew & Dickerson, 1981; Feuerstein et al., 1986). They also are known to cause a conformational transition of B- to Z-form DNA in poly[d(G-C)] (Behe & Felsenfeld, 1981; Chen et al., 1984). In order to examine the effects of spermidine on various DNA structures in the *Adh* gene region, spermidine was added to the supercoiled psAc2 DNA at various

Table 3.2.1
List of nucleases and protein for detection of altered DNA structures

nuclease or protein (abbreviation)	souce	product[a]	pH in reaction	specific features[b] examined
S1 nuclease (S1)	*Aspergillus oryzae*	ds	4.5	cruciform (7, 10) Z-DNA (14) poly(pu):poly(py) (1, 12) AT-rich (10)
mung bean nuclease (MBNase)	mung bean spout	ss	7.0	AT-rich (13)
N. crassa nuclease (NCNase)	*Neurospora crassa*	ds	7.9	poly(pu):poly(py) (15)
Bal 31 nuclease (Bal 31)	*Altermonas espejiani*	ds	8.0	Z-DNA (5) cruciform (8)
phospho-diesterase (PDE I)	*Crotalus adamanteus*	ds	9.2	———— (11)
T4 endonuclease VII (endo VII)	T4 phage-infected *E. coli*	ds	8.0	Holliday structure (9, 4) cruciform (3, 8)
single strand binding protein (SSB)	*E. coli*	—	7.4	single stranded DNA (6, 2)

[a] ds; double strand cut, ss; single strand cut
[b] Data are derived from publications cited in parenthesis;
1. Canton & Efstratiadis (1983), 2. Gelikin et al. (1983), 3. Greaves et al. (1985), 4. Jensch & Kemper (1986), 5. Kilpatrick et al. (1983), 6. Kowalczykowski et al. (1981), 7. Lilley (1980), 8. Lilley & Hallam (1984), 9. Mizuuchi et al. (1982), 10. Panayotatos & Wells (1981), 11. Prichard et al. (1977), 12. Schon et al. (1983), 13. Shelfin & Kowalski (1984), 14. Singleton et al. (1983), 15. Wang & Hogan (1985)

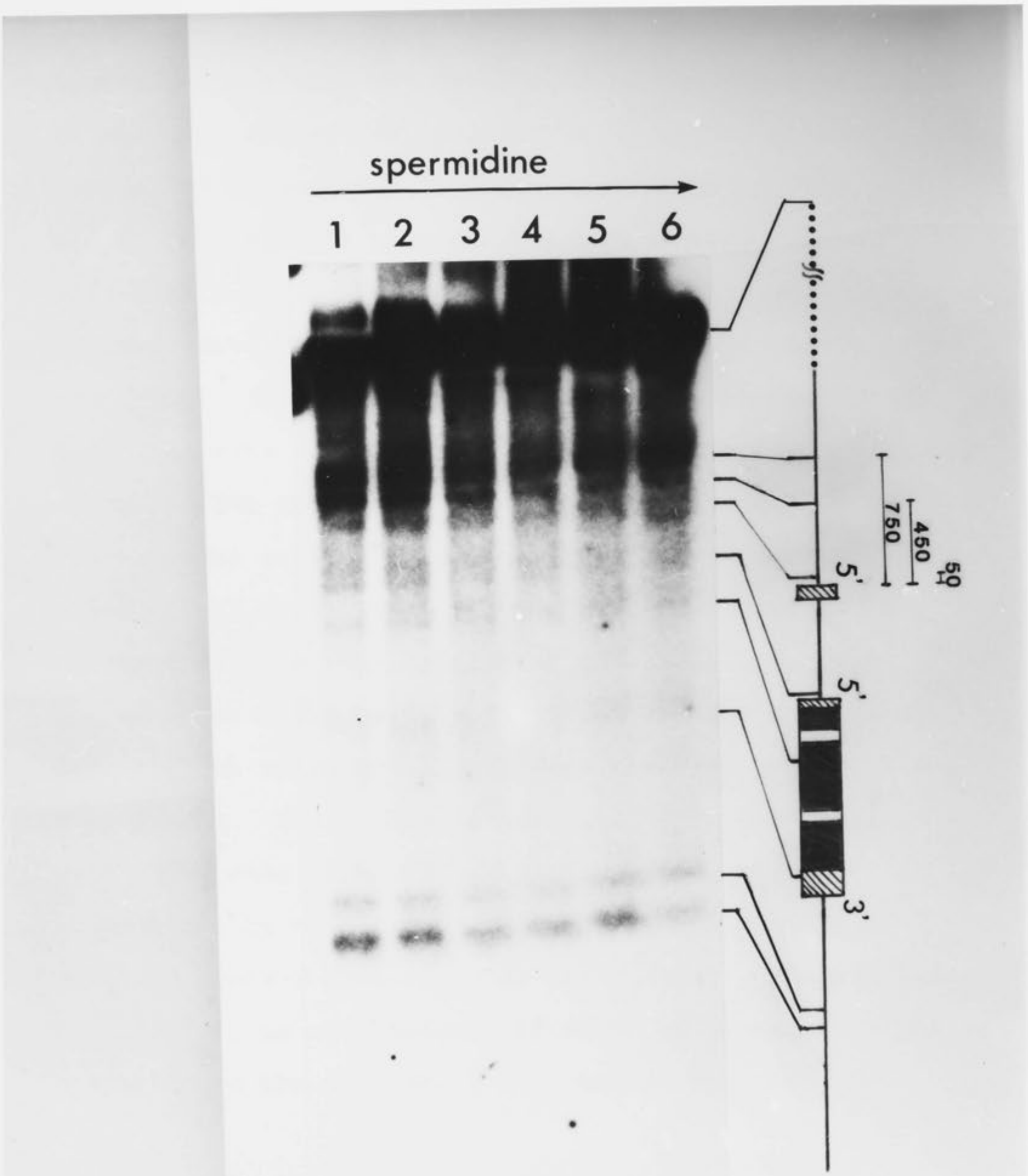
concentrations (0 to 5 mM), followed by S1 nuclease digestion, and then analysed as above. The results were shown in Fig. 3.2.8. Spermidine did not lead to any obvious alteration of the pattern of S1 hypersensitive sites on the DNA. It was noticed, however, that the overall susceptibility of the DNA to S1 nuclease slightly decreased on the addition of spermidine. Such an inhibitory effect of spermidine is in accordance with other reports (Srivenugopal & Morris, 1985). This was likely to be due to the aggregation of DNA with the spermidine.

3.2.7. Examination of altered DNA structures with other nucleases

I also examined the altered DNA structures of the *Adh* gene region in the physiologically supercoiled DNA, psAc2, by using various different nucleases, i.e., mung bean nuclease (MBNase), phosphodiesterase I (PDE I), Bal 31 nuclease (Bal 31), T4 endonuclease VII (endo VII) and *N. Crassa* nuclease (NCNase). These nucleases have their own specificities under different conditions as briefly described in materials and methods (Chapter 3.2.2, see also Table 3.2.1). In the previous experiments, the DNA of the *Adh* gene region was incubated with S1 nuclease in an acidic buffer (pH 4.6). Here, however, the pH values of buffers used for MBNase, PDE I, Bal 31, endo VII and NCNase are 7.0, 9.2, 8.0, 8.0 and 7.0 respectively. Also, MBNase produces only nicked circular DNA by single strand cut rather than linearised DNA, while the other nucleases produce linearised DNA by double strands cut similar to S1

Fig. 3.2.8 Effects of spermidine on the distribution of S1 nuclease hypersensitive sites at the *Adh* gene locus.

The supercoiled *Adh* DNA, psAc2, was incubated at 30°C for 10 min with various concentrations of spermidine; 0 (lane 1), 0.25 (lane 2), 0.5 (lane 3), 1.0 (lane 4), 2.5 (lane 5), 5.0 (lane 6) mM . Then the DNA was digested at 30°C for 10 min with 125 units/ml S1 nuclease. The S1 hypersensitive sites were detected as described in the legend to Fig. 3.2.4.


spermidine
1 2 3 4 5 6
5'
5'
3'
750
450
50

nuclease. Therefore, to detect sites nicked by MBNase, the MBNase treated DNA was further digested with S1 nuclease, which preferentially cut the opposite strand at a nicked site (see Chapter 3.2.2 for detail).

First, by digestion with MBNase, a major hypersensitive site was detectable in the *Adh* gene region in the supercoiled DNA, psAc2 (Fig. 3.2.9). This site was mapped 750 bp upstream from the adult capping site, where one of the major S1 hypersensitive site also existed (see also Fig. 3.2.1). Although some other sites were detected in some cases, they were not constantly apparent or were too faint. They were attributed possibly to the second digestion with S1 nuclease. To analyse the conformational change of DNA in the *Adh* gene region under various degrees of torsional stress, psAc2-topoisomers with various superhelical densities were treated with MBNase. The susceptibility of the major site to MBNase gradually decreased as the superhelical density became lower (Fig. 3.2.10). No other sites were detectable by MBNase on any topoisomers. These observations were in sharp contrast with those when using S1 nuclease. While, the susceptibility of the major site hypersensitive to S1 nuclease rapidly diminished as the density decreased (see Fig. 3.2.2), no new hypersensitive sites appeared on MBNase digestion on the moderately supercoiled DNA.

Secondly, in digestion with PDE I, only one major hypersensitive site was detected. The site was mapped slightly downstream from the major site of MBNase or 550 bp

Fig. 3.2.9 Distribution of mung bean nuclease hypersensitive sites at the *Adh* locus.

The supercoiled *Adh* DNA, psAc2, was digested at 30°C for 20 min with 25 units/ml. The nicked circular DNA produced by MBNase was completely cleaved by EcoRI to be linearized. The linearized DNA was further digested with S1 nuclease at various concentrations of 0 (lane 2), 67 (lane 3), 167 (lane 4), 333 (lane 5) and 660 units/ml (lane 6). The nicked circular DNA produced by MBNase which was not further treated with any enzymes (lane 1) and the intact psAc2 DNA cleaved by EcoRI and digested with 660 units/ml of S1 nuclease without prior MBNase treatment (lane 7) were used as controls. The DNAs were electrophoresed on 1% agarose gel followed by indirect end-labeling methods. The PvuI-EcoRI fragment was used as a probe.

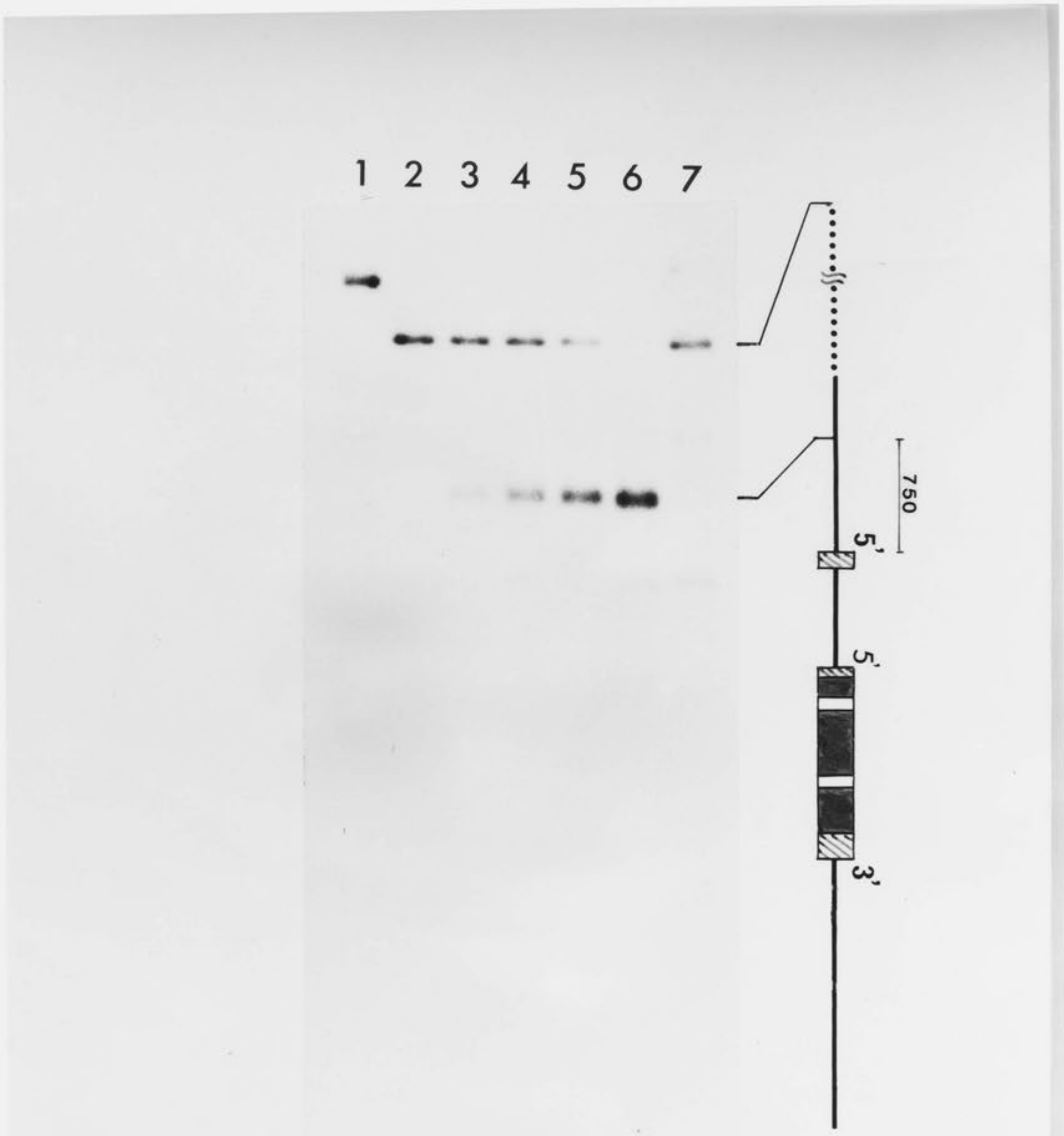
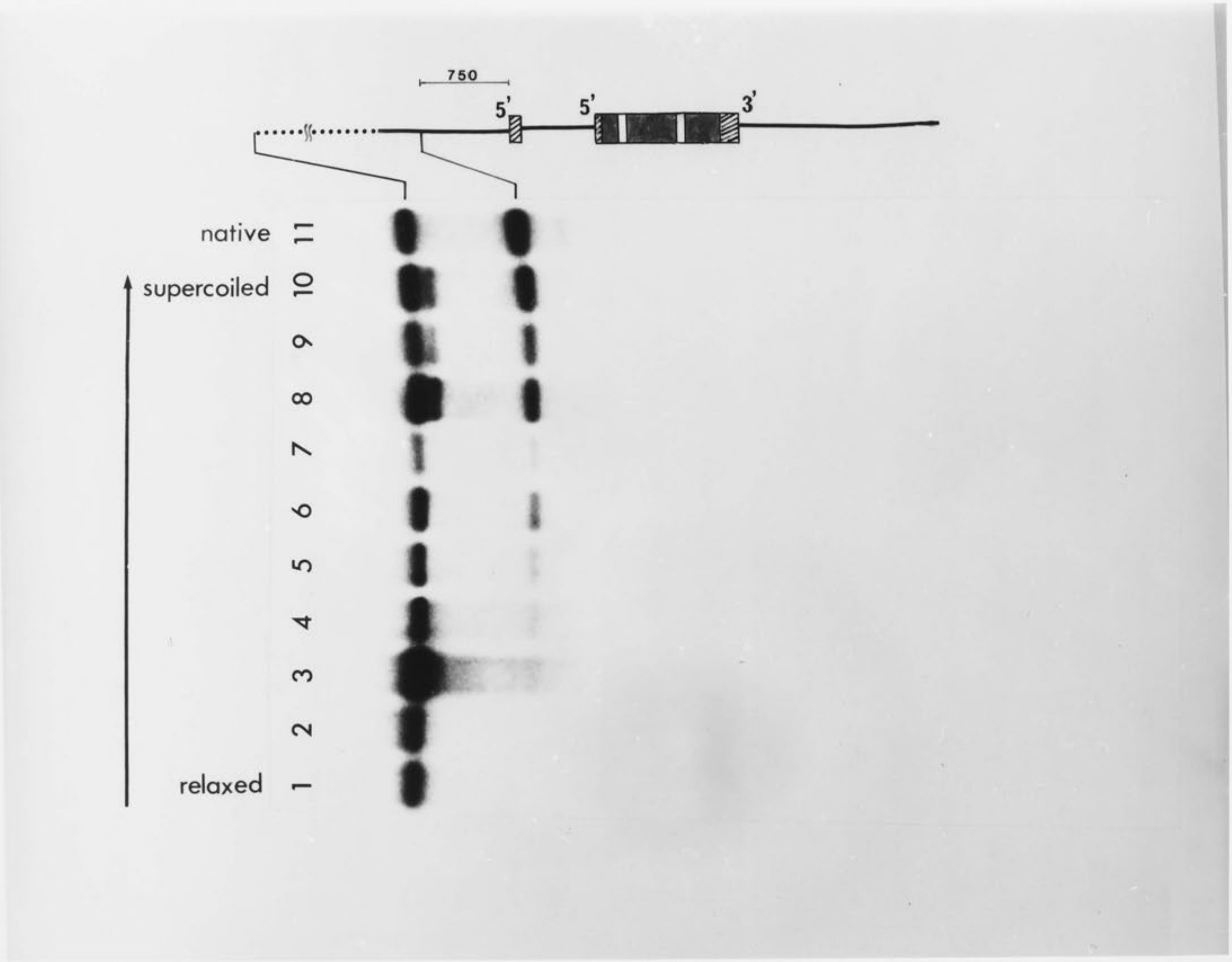

1 2 3 4 5 6 7
750
5'
5'
3'

Fig. 3.2.10 Dependence of mung bean nuclease sensitivity of a specific site at the *Adh* locus on superhelical densities of topoisomers.

The psAc2-topoisomers used in this experiment were the same to those in Fig. 3.2.6. The fragments were digested 30°C for 20 min with 2.5 units/ml mung bean nuclease, and cleaved by EcoRI. The linearized DNA was purified and further digested 30°C with 333 units/ml S1 nuclease. The MBNase hypersensitive sites were detected as described in the legend to Fig. 3.2.9.


750
5'
5'
3'
native
supercoiled
relaxed
11
10
9
8
7
6
5
4
3
2
1

from the adult capping site (Fig. 3.2.11). The digestion buffer for PDE I was very alkaline; pH 9.2. Therefore, this observation suggests that the region 700-800 bp upstream has characteristic DNA sequences for the formation of altered structures in both acidic and alkaline solutions. Differences in the nuclease specificity or in pH may lead the recognition site to shift slightly, and this site may be related to a major site hypersensitive to S1 nuclease.

Thirdly, sites hypersensitive to Bal 31 were detected on the supercoiled DNA (Fig. 3.2.12). Of particular interest was that the pattern of sites hypersensitive to Bal 31 was almost identical to that of S1 nuclease, though the susceptibility of the site 750 bp upstream from the adult capping site was relatively high and an additional site was detectable 550 bp upstream from the larval capping site. Therefore, the overall pattern produced by Bal 31 showed a similarity to that of S1 nuclease in the presence of 20 mM of Mg(II) ion (see also Fig. 3.2.7).

The different patterns of hypersensitive sites observed with various types of nucleases may in part be due to the delicate difference in their recognition specificities and to the varied digestion conditions. However, it was at least certain that since some of the sites were detectable by various types of molecules at different pH values, altered DNA structures were constantly formed at these sites under torsional stress over a wide range of pH. In particular, an altered DNA structure was

Fig. 3.2.11 Phosphodiesterase I hypersensitive site at the *Adh* locus on the supercoiled DNA.

The supercoiled *Adh* DNA, psAc2, was digested at 30°C for 20 min with 0.0125 units/ml phosphodiesterase I. The DNA was further digested with EcoRI and electrophresed on 1% agarose gel followed by indirect end-labeling methods (PDEI). To compare the locations of hypersensitive sites, the results of digestion by mung bean nuclease (MBNase) and S1 nuclease (S1) are also shown.

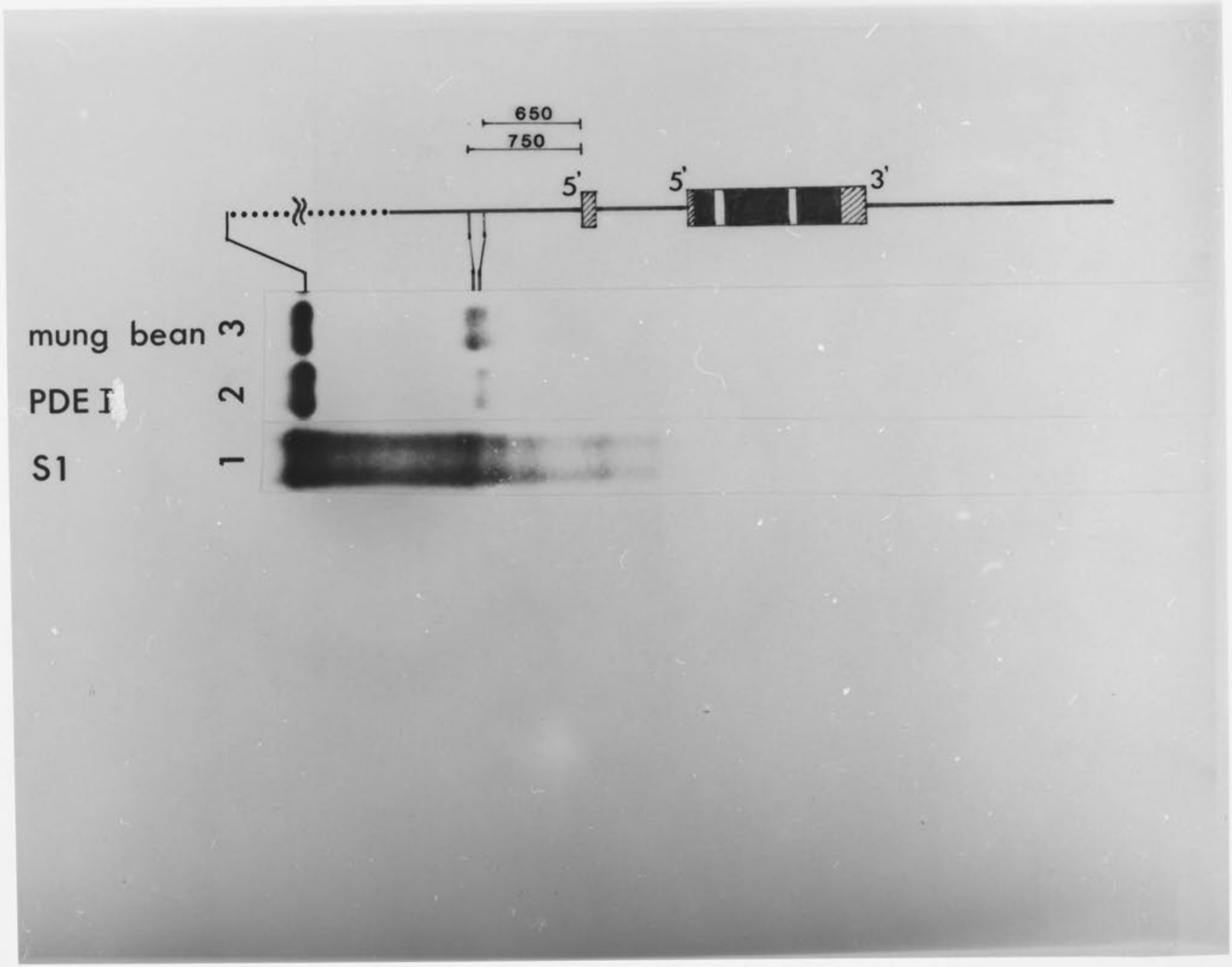
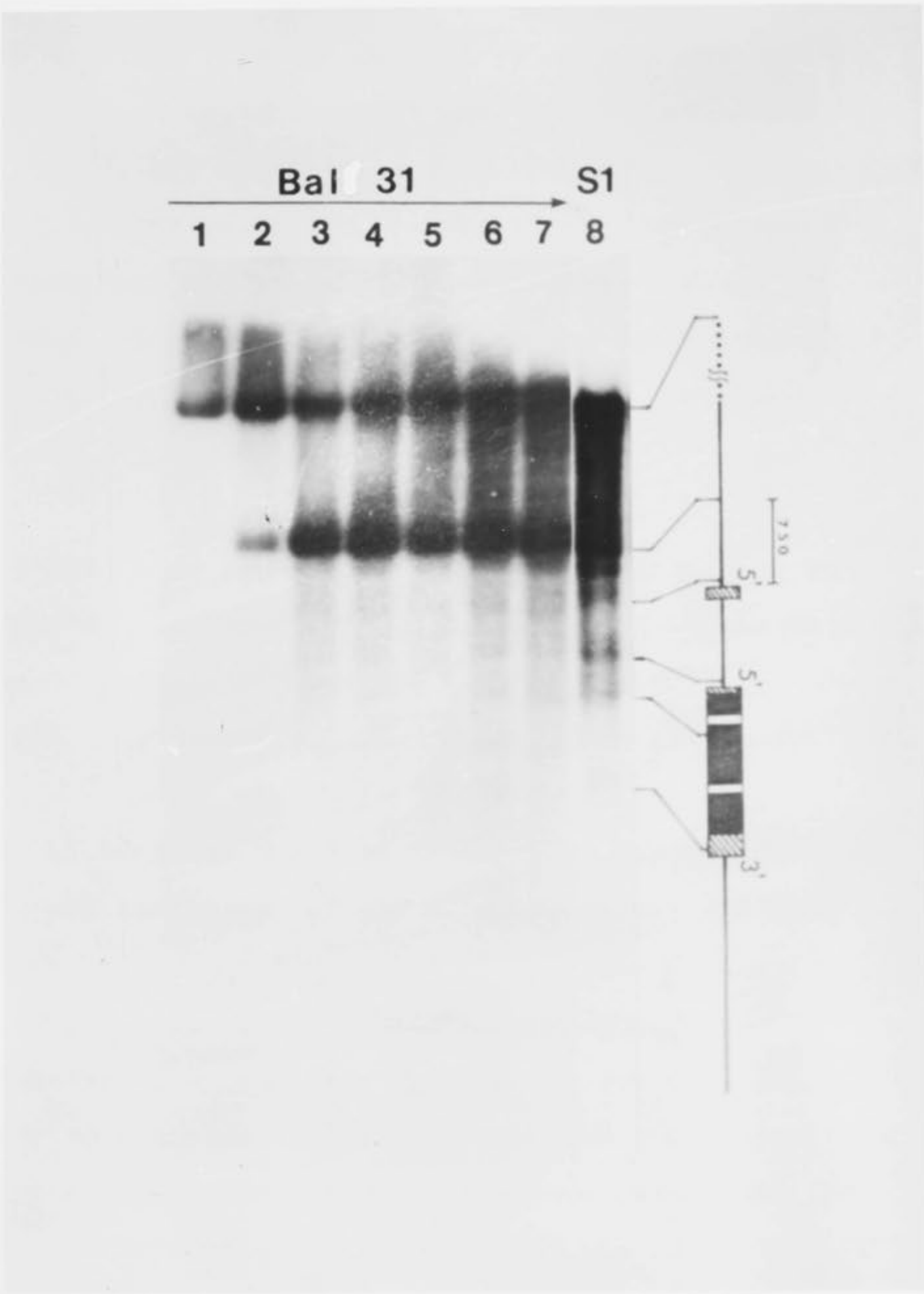

650
750
5'
5'
3'
mung bean
PDE I
S1
3
2
1

Fig. 3.2.12 Distribution of Bal 31 nuclease hypersensitive sites at the *Adh* locus on supercoiled DNA.

The supercoiled *Adh* DNA, psAc2, was digested at 30°C for 20 min with Bal 31 nuclease at various concentrations; 0 (lane 1), 0.0125 (lane 2), 0.025 (lane 3), 0.05 (lane 4), 0.125 (lane 5), 0.25 (lane 6) and 0.5 units/ml (lane 7). The digested DNA was cleaved by EcoRI, electrophoresed on 1% agarose gel, followed by indirect end-labeling methods. The PvuI-EcoRI fragment was used as a probe. The results of digestion by S1 nuclease are also shown in lane 8.

extremely stable around 750 bp upstream from the adult capping site in solution of any pH tested.



extremely stable around 750 bp upstream from the adult capping site in solution of any pH tested.

Finally, I also examined sites hypersensitive to T4 endonuclease VII on the *Adh* gene region in the supercoiled psAc2 DNA (Fig. 3.2.13). Endo VII specificity recognises 'Holliday structure' or a branched strand region of DNA (Mizuuchi et al., 1982; Jensch & Kemper, 1986). The site 750 bp upstream from the adult capping site was extremely susceptible to digestion with endo VII. Furthermore, two other sites, 550 bp upstream from the larval capping site and 550 bp downstream the 3'- end of the *Adh* gene, were also detected, but rather faintly. These observations suggested that at least at the site 750 bp upstream from the adult capping site, a stable Holliday structure was formed.

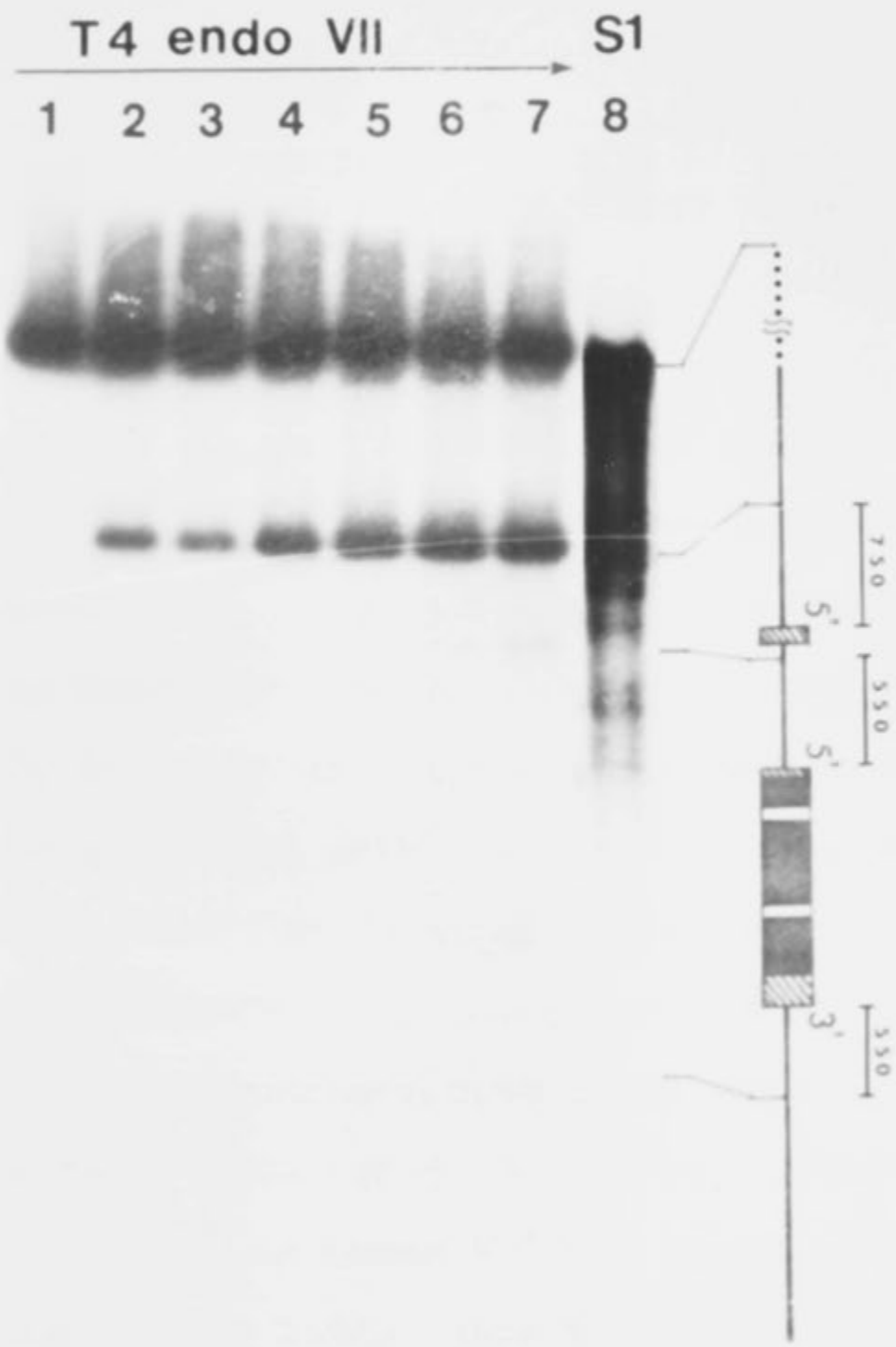
I also used the *N. crassa* single strand specific nuclease for the detection of altered DNA structures in the *Adh* gene region. However, no discrete sites were detectable by digestion with NCNase (data not shown).

3.2.8 S1 nuclease hypersensitive sites further upstream and downstream of the *D. melanogaster Adh* gene

Calculating the hypothetical borders of the territories of a gene (see Chapter 3.1 for details), I considered the possibility that particular conformational properties may be involved in DNA sequences around the borders of intrinsic territories. In the case of the *D.*

Fig. 3.2.13 Distribution of T4 endonuclease VII hypersensitive sites at the *Adh* locus on supercoiled DNA.

The supercoiled *Adh* DNA, psAc2, was digested at 30°C for 20 min with T4 endonuclease VII at various concentrations; 0 (lane 1), 25 (lane 2), 50 (lane 3), 100 (lane 4), 250 (lane 5), 500 (lane 6) and 1,000 units/ml (lane 7). The digested DNA was cleaved by EcoRI, electrophoresed on 1% agarose gel, followed by indirect end-labeling methods. The PvuI-EcoRI fragment was used as a probe. The results of digestion by S1 nuclease are also shown in lane 8.


T4 endo VII
S1
1 2 3 4 5 6 7 8
750
550
550
5'
3'

melanogaster Adh gene, it is thought that while the larval transcriptional unit possess 0.9 kb long intrinsic territories on both flanking regions, the adult transcriptional unit possesses 2.9 kb long intrinsic territories.

The previous investigations concentrated upon S1 nuclease hypersensitive site present about 1 kb upstream and downstream from the *Adh* gene. In this experiment, to characterise the intrinsic territories of both transcriptional units, I analysed altered DNA structures under torsional stress in both flanking regions up to 4 or 5 kb away from the *Adh* gene using a recombinant DNA with a large insertion, psAF2 (Goldberg, 1980; Fig. 3.2.14 A).

The physiologically supercoiled psAF2 DNA was incubated with S1 nuclease, restricted with Bam HI enzyme, then S1 hypersensitive sites were detected by indirect end labelling methods as described above. I used two short DNA fragments (about 200 bp long) as hybridization probes for detection of S1 hypersensitive sites in the upstream and downstream regions from the Bam HI site, respectively (illustrated in Fig. 3.2.14 A). By using the Bam HI-upstream probe, two major hypersensitive sites were detected in the upstream region (Fig. 3.2.14 B). One was mapped 2.7 kb upstream and the other 0.75 kb upstream from the adult capping site. The latter site was apparently identical to that detected on the smaller cloned DNAs, psAc1 and 2, in the previous experiments (see Chapters 3.2.3 to 3.2.7). The intensities of these two bands on the

autoradiograph allow us to compare the relative susceptibilities of these sites. The 2.7 kb upstream site was more susceptible to digestion with S1 nuclease than the 0.75 kb upstream site.

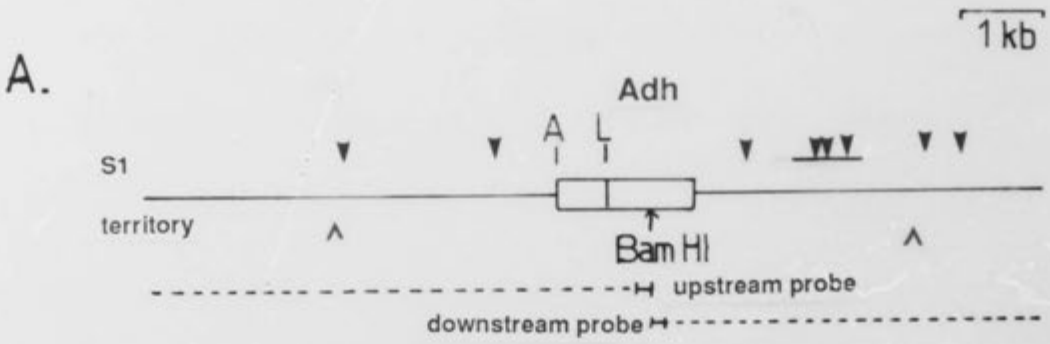
The blotting filter used for detection in the upstream region was washed out to remove the Bam HI-upstream probe, then the filter was re-hybridized with the Bam HI-downstream probe. The location of S1 hypersensitive sites was given by the size of fragments downstream from the Bam HI site. In the downstream region of the *Adh* gene, several S1 hypersensitive sites were detected (Fig. 3.2.14 C). Three major hypersensitive sites are mapped approximately 3.0, 1.6 and 0.8 kb downstream from the 3′- end of the *Adh* gene. Hypersensitive sites detected in a higher part of the autoradiograph were located in the pBR322 vector region. Of particular interest was that S1 hypersensitive sites were clustered in a 1.3 - 2.2 kb region downstream from the 3′- end of the *Adh* gene. It was noticed once again that the 0.8 kb downstream site was identical to that detected on the psAc 1 and 2 DNAs. In the linearised psAF2 DNA, no S1 hypersensitive sites were detectable either upstream or downstream from the *Adh* gene (data not shown).

From these observations, it was apparent that some particular DNA structures sensitive to S1 nuclease under torsional stress are indeed formed at the sites where I had predicted the borders of the intrinsic territories of both *Adh* transcriptional units should be found. Namely, the 2.7 kb upstream and 3.0 kb downstream sites seem to correspond

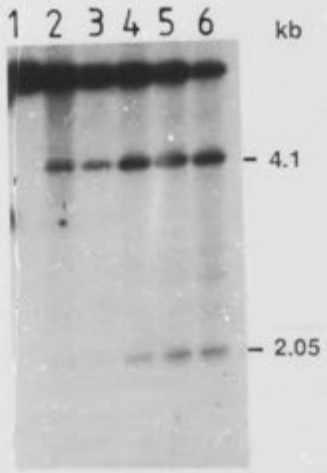
Fig. 3.2.14 S1 hypersensitive sites far upstream and downstream from the *Adh* gene on supercoiled DNA.

(A) The map of the 11.8 kb SacI fragment of the *Adh* locus, which is cloned in pBR322 vector (Goldberg, 1980), is schematically shown. Two probes used in this experiment (BamHI upstream and downstream) are indicated below the map. The arrows indicate the positions of S1 hypersensitive sites in this experiment. (B) The supercoiled *Adh* DNA, psAF2, was digested at 30°C for 20 min with S1 nuclease at various concentrations of 0 (lane 1), 50 (lane 2), 125 (lane 3), 250 (lane 4), 500 (lane 5) and 1,000 units/ml (lane 6). The DNA was further cleaved by BamHI, and then electrophoresed on 1% agarose gel followed by indirect end-labeling methods. The BamHI-upstream probe was used as a hybridization probe to detect S1 hypersensitive sites far upstream from the *Adh* gene. (C) The hybridization filter used in (B) was washed to remove the previous probe, then re-hybridized with BamHI-downstream probe to detect S1 hypersensitive sites far downstream from the *Adh* gene.

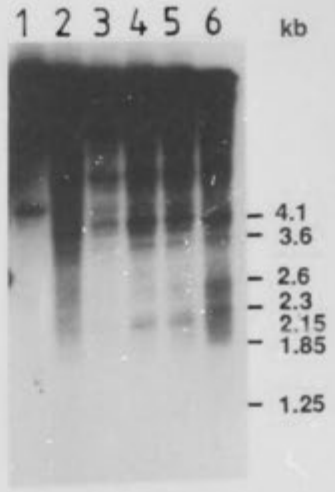
A.


1 kb
Adh
A
L
S1
territory
BamHI
upstream probe
downstream probe

B.


1 2 3 4 5 6
kb
4.1
2.05

C.


1 2 3 4 5 6
kb
4.1
3.6
2.6
2.3
2.15
1.85
1.25

to the demarcation points for the upstream and downstream territories of the adult transcriptional unit (1.8 kb long), respectively. Furthermore, the 0.75 kb upstream and 0.8 kb downstream sites also seem to correspond to those of the larval transcriptional unit (1.1 kb long), respectively.

3.2.9 Discussion

The experiments described in this chapter 3.2 clearly demonstrated the plasticity of the DNA molecule in the *D. melanogaster Adh* gene region when this molecule is under torsional stress. This plasticity was detected as an alteration of DNA structures formed around the *Adh* gene by using several kinds of nucleases and other reagents. The location of these plastic DNA structures around the *Adh* gene is schematically shown in Fig. 3.2.15.

There are several features of the DNA structures to be discussed in the context of biological significance. First, digestion of the superhelical topoisomers with S1 nuclease showed that various S1 hypersensitive sites were differentially detected among the topoisomers. This result provides us with evidence that DNA structures are dynamically changeable depending upon the strength of torsional stress. Thus, DNA molecules are conformationally polymorphic. Furthermore, it suggests that conformational transitions on distinct sites result from their mutual interaction over a long distance throughout the same topological domain of a circular DNA molecule. Recently,

Fig. 3.2.15 **Positions of altered DNA structures detected at the *Adh* locus by various nucleases and reagents.**

Vertical arrows indicate the hypersensitive sites detected on the supercoiled DNA in every case. Arrowheads indicate the S1 hypersensitive sites detected on moderately supercoiled DNA or in the presence of low concentrations of SSB. Larger arrows or arrowheads represent higher sensitivities at the sites.


0
0.5
1.0 Kb
adult
larvae
5'
5'
3'
S1
+ Mg2+
s.s.b.
mung bean
phosphodiesterase
endonuclease VII
Bal 31

Kelleher III et al. (1986) have reported that in two discrete B-Z transitions, one conformational transition in topologically constrained duplex DNA is affected by the other transitional process. Thus, conformational transitions induced by torsional stress in a topological domain compete with each other for the free energy of supercoiling.

Secondly, I have demonstrated that the microenvironment of a DNA molecule in solution affects the formation of altered DNA structures at different sites in various manners. For example, Mg(II) ion and pH can readily change the DNA conformation. It was found that even under a given torsional stress, DNA structures in solution were not fixed tightly but were still dynamically changeable by various effects. However, KCl and spermidine at various concentrations did not exert any significant influence on the DNA conformation. It seems likely, that in a supercoiled DNA rather than a linear one, the superhelical restraint prevents any local conformational transitions being induced by these reagents.

Thirdly, some of the sites with altered DNA structures under torsional stress correspond to biologically functional sites of the *Adh* gene, i.e., the adult and larval promoter sites, the second intron and poly(A) addition site. Therefore, the altered DNA structures at these sites may have functions in the respective processes comprising transcription of the *Adh* gene. The 5'- flanking regions of both adult and larval capping sites in

particular possess elastic properties with respect to DNA conformation. In these two regions, DNA conformation is readily changed by various factors, e.g., the strength of torsional stress, the concentrations of Mg(II) ion and SSB. These observations raise the possibility that all these factors are involved in regulation mechanisms during transcriptional activity of a gene *in vivo*. The idea seems attractive to me that the 5'- flanking regions of both capping sites play an important role in the regulation of differential transcription of the adult and larval units of the *Adh* gene through the plastic properties of the DNA molecule.

Moreover, it is apparent that there are several types of altered DNA structures in the *Adh* gene region. The altered DNA structure 750 bp upstream from the adult capping site was detectable by various kinds of nucleases under varied conditions. The structure was readily recognised by SSB, and the recognised region seemed to extend unidirectionally upstream from the site according to the concentration of SSB.

In contrast, the altered DNA structures formed at other sites in the *Adh* gene region were not detectable by all of the used nucleases. Perhaps different behaviours among these sites may result from the fine differences in the features of altered DNA structures which can be differentially recognised by individual nucleases under specific conditions.

Finally, it should be emphasized here that the demarcation points of intrinsic territories for the adult and larval transcriptional units are marked by altered DNA structures, which are detectable under torsional stress by S1 nuclease digestion. As described previously (see Chapter 3.1), in many other genes, such altered DNA structures were also detected as demarcation points around the borders on intrinsic territories. Since gene territories are assumably required for active functioning of a gene, it seems likely that altered DNA structures found at demarcation sites are intimately involved in mechanisms regarding territorial effects. Furthermore, SSB seems to distinguish between the coding sequence of the *Adh* gene and the non-coding sequences. Thus, DNA sequences may be organised around a gene in a particular orderly manner and contain information to build a specific chromatin structure.

Plastic DNA structures which can dynamically alter according to the superhelical density and ionic conditions, may play important roles in gene functions. In many other genes, further experiments upon plastic structures would lead us to better understanding of gene regulation in terms of conformation.

3.3 DNA sequences of the *D. melanogaster Adh* gene and fine mapping of altered DNA structure sites

3.3.1 Introduction

Negative supercoiling of duplex DNA has been known to promote the conformational transition from B-form structure to altered DNA structure at local regions. These regions possess specific nucleotide sequences which have been characterised in some cases. First, alternating d(CG)·d(AT) sequences can adopt the left-handed helical conformation (Z-form) in negatively supercoiled DNA (reviewed by Rich et al., 1984). In this case, S1 nuclease recognises aberrant structural features at the junction between neighbouring right- and left-handed DNA regions (Singleton et al., 1982).

Secondly, inverted repeat sequences can also adopt cruciform geometry in which complementary sequences on the same strand form intramolecular double stranded structures (Lilley, 1980; Panayotatos & Wells, 1981). The loops and stems of cruciform structures are specifically cleaved by S1 nuclease and chemical reagents (Lilley, 1980; Soholten & Nordheim, 1986). Similarly, d(AT)n·d(AT)n tracts can also form a cruciform structure, which is recognised by S1 nuclease and T4 endonuclease VII (Greaves et al., 1985; Haniford & Pulleybank, 1985).

Thirdly, short direct repeat sequences, d(GA)n, which may allow a slippage structure, are also susceptible to

digestion with S1 nuclease (Mace et al., 1983; Hentoschel, 1982; Htun et al., 1984). Fourthly, homopurine-homopyrimidine sequences are also susceptible to the digestion with S1 nuclease (Schon et al., 1983; Nickol & Felsenfeld, 1983). The possibility that these sequences become single stranded under torsional stress was examined by using bromoacetaldehyde, which reacts with single stranded nucleotides of a DNA duplex (Kohwi-Shigematsu et al., 1983; Kohwi-Shigematsu & Kohwi, 1985). In spite of their susceptibility to S1 nuclease and bromoacetaldehyde, however, doubts about the single strandedness in such sequences were raised by the results obtained with dimethylsulfonate treatment and by thermodynamic considerations (Canton & Efstratiadis, 1984). Further, the protonation at poly d(TC)·d(GA) was detectable under torsional stress in pH dependent manner. This suggested that the sequences are either Hoogsteen or reverse Hoogsteen base-paired to protonated dC residues of polypyrimidine strands, and that normal Watson-Crick base pairing is not involved there (Pulleybank et al., 1985). So far, although altered DNA structures or non-B-form structures are certainly formed at these sequences under torsional stress, the actual structural detail still remains unclear.

The previous experiments (Chapter 3.2) showed that altered DNA structures were formed at specific sites on the *Adh* gene region under torsional stress. Are the known structures listed above, e.g. cruciform, formed at these sites? What are the specific nucleotide sequences which

are responsible for the dynamic alteration of DNA conformation observed on the *Adh* gene region? To answer these questions, I have determined the nucleotide sequences which range 383 to 1095 bp upstream from the adult capping site of the *Adh* gene. I have analysed the free energy distribution of DNA duplex around the *Adh* gene. Furthermore, the major S1 hypersensitive sites were mapped at nucleotide level.

3.3.2 Materials and methods

(a) Nucleotide sequencing

0.3 kb Eco RI - Eco RI, 0.7 kb Nde I - Nde I, 0.4 kb Nde I - Xba I and 0.3 kb Xba I - Nde I fragments upstream from the *Adh* gene were isolated from the *Adh* recombinant DNA, psAc1, and subcloned into pUC 13 and M13. These DNA fragments were sequenced by the dideoxyribonucleotide method (Hattori & Sakaki, 1986; Messing, 1983) and determined from both strands. Furthermore, the sequences about 150 bp upstream and downstream the Xba I site were determined by Maxam and Gilbert method (1980) and confirmed.

(b) Computer analysis of nucleotide sequences

Sequence analysis was performed using a Dec 10 system computer. Sequences were compared using NUC ALN (Wilbur & Lipman, 1983) and SEQ (Brutlag et al., 1982). The values of the free energy of nearest-neighbour interactions were

obtained by Breslauer et al. (1986). To analyse thermodynamic stability of DNA duplex, the values of free energy were plotted against positions of nearest-neighbour interactions. The programming for graphing the free energy was performed by Mr. D. Smith (R.S.B.S., A.N.U., Australia).

(c) Fine mapping of altered DNA structures

The *Adh* recombinant plasmid DNAs, psAc1 and psAF2, were partially digested to produce more than one third DNA as nicked-circular with S1, mung bean nucleases and T4 endonuclease VII as described previously (see Chapter 3.2.2). The nuclease-digested DNA was restricted with an appropriate enzyme, e.g. Eco RI, Hind III and Xba I. These restricted sites were labelled at either 3′- or 5′- end (Maxam & Gilbert, 1980). The labelled DNA was further restricted with another enzyme, and run on 1% agarose gel, and appropriate fragments were isolated as previously described (see Chapter 2).

As sequence makers, the same DNA fragments as above were obtained from the nuclease-untreated DNAs, and treated to cut the DNA strand preferentially at adenosine to cytosine (Maxam & Gilbert, 1980). The end-labelled DNA fragments was analysed on 6% sequencing gels together with a sequence marker.

3.3.3 Nucleotide sequences upstream of the *D. melanogaster* *Adh* gene

The nucleotide sequences of 0.7 kb Nde I and 0.3 kb Eco RI fragments upstream of the *D. melanogaster Adh* gene were determined as described in Materials and methods. The nucleotide sequence 5'flanking to the *Adh* gene is shown in Fig. 3.3.1, which includes previously determined sequence (Kreitman, 1983). The nucleotide sequence of the 0.3 kb Eco RI fragment is also shown in Fig. 3.3.1.

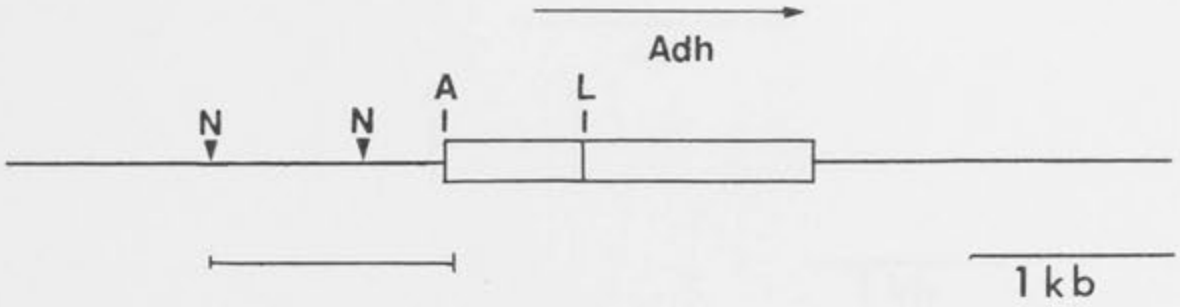
I have compared the nucleotide sequences of the 5'-flanking regions and the untranslated leader region of the adult and larval *Adh* units. The results showed that there was no significant homology between them (data not shown). Not even short homologous elements were found.

Evidence has accumulated that quite wide regions around the 5'- and 3'- flanking regions of the genes present in a gene cluster are conserved as well as the coding regions of these genes. Good examples are β- or β-like globin genes of various species of organisms (Efstratiadis et al., 1980), the *Drosophila glue* gene (Garfinkel et al., 1983) and the *Drosophila hsp* 70 genes (Mason et al., 1982). The sequence homologies of these genes implicate gene conversion and rearrangement during evolution. In contrast, the slight homology between the 5'- flanking regions and the untranslated regions of the *Adh* gene does not suggest that the dual promoter gene was produced by a simple duplication of the 5'-region of a proto-*Adh* gene.

Fig. 3.3.1 Nucleotide sequences of the regions upstream from the *D. melanogaster Adh* gene.

The 1,195 bp-nucleotide sequence in the 5'-flanking region to the adult capping site (A), and the 277 bp-nucleotide sequence in the EcoRI fragment about 2.7 kb upstream from the adult capping site (B) are shown. The nucleotide sequences in the NdeI and EcoRI fragments were determined as described in Chapter 3.3.2. Other sequences shown in (A) are reproduced from Kreiteman's data (1983). The adult capping site and the transcriptional direction are shown by a horizontal arrow. TATA-box is boxed. (The major restriction sites used in the following experiments are underlined.) A 12-nucleotide direct repeat is indicated by horizontal arrows under the sequence (see also Fig. 3.3.4). Vertical arrows above and below the sequences represent S1 cleavage sites on the sense and anti-sense strands, respectively, shown in Fig. 3.3.5, 6, 7 and 8. Arrows with asterisk represent endonuclease VII cleavage sites. The sizes of arrows reflect the relative frequency of cleavage. The schematic map of the *Adh* locus. The regions of the nucleotide sequences shown in (A) and (B) are schematically shown above the respective sequences. N and E indicate Nde I and Eco RI sites, respectively. A and L indicate the initiation sites of the adult and larval transcripts of the *Adh* gene, respectively.

A



CTATGAATTTGCAGTAGCCTCCTCCCACATCATATTTAACGCCCTATATTCAAAATTTGCTCAAGAAAATATTTGAACCAAATTGATTTTTAGTC
NdeI

-1000
AATTAGTTTTTAAGTAATTAAGTGGAGTAAACATATACAATTTTATTCTTACCAAACACATATACTCATATATTTTGAATAAATAAATAAACAAATATAT

-900
ATAAAATCTACGAAATTGGCAAACAAATTTTAAAGCATTATAGTATTGCCGATTTAATTAATATAATTAAATAATATGTACATGTATTAATCTTGTGTGC

-800
GAGCATGGGTTAAATCTAGCTGCATTCGAAACCGCTACTCTGGCTCGGCCACAAAGTGGGCTTGGTCGCTGTTGCGGACAAGTGAGATTGCTAATGAGCT

-700
GCTTTTAGGGGGCGTGTTGTGCTTGCTTTCCAACTTTTCTAGATTGATTCTACGCTGCCTCCAGCAGCCACCCCTCCCATCACCATCCCCATCACCATCC
XbaI

-600
AGTCCCGTTGGCTCCCAGTCACAGTATTACACGTATGCAAATTAAGCCGAAGTTCAATTGCGACCGCAGCAACAACACGATCTTTCTACACTTCTCCTTG

-500
CTATGCTTGACATTCACAAGGTCAAAGCTCTTAATATTCTGGCTCGTGGCCCTACACTGTAAGAAATTACTATAGAAATAACGGTACACGGAATAAGATA

-400
TTTTTTTTAGTCCATATGCTTTTAACAAATGTGTTTTGAGTTTATGTTATATTATTGTTAGAAAACAGGTGTTTTTTTTTAAATCGGTTAAAAAATTACT
NdeI

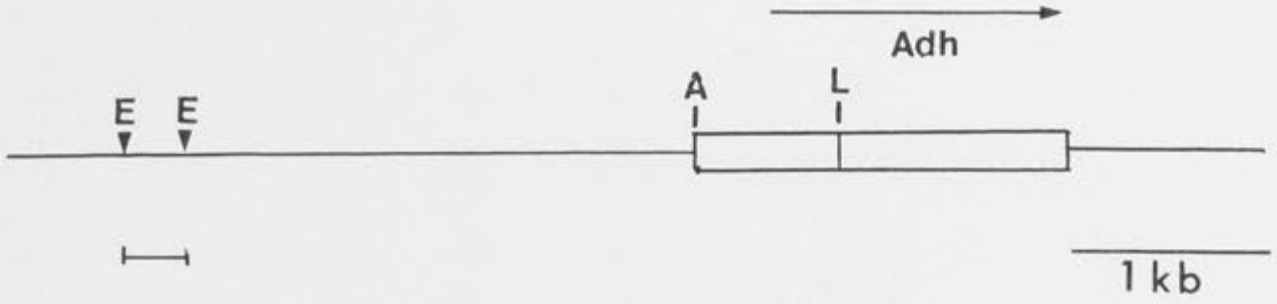
-300
ACGAGAGAAAAATACAAATTTTGTAAATAAGATTGACTCTTTTTCGATTTTGGAATATTTTCATTCATTTTATGTTTTTACGTTTTCACTTATTTGTTTC

-200
TCAGTGCACTTTCTGGTGTTCCATTTTCTATTGGGCTCTTTACCCCGCATTTGTTTGCAGATCACTTGCTTGCGCATTTTTATTGCATTTTACATATTAC

-100 -1
ACATTATTTGAACGCCGCTGCTGCTGCATCCGTCGACGTCGACTGCACTCGCCCCCACGAGAGAACAGTATTTAAGGAGCTGCGAAGGTCCAAGTCACCG

+1
ATTATTGTCTCAGTGCAGTTGTCAGTTGCAGTTCAGCAGACGGGCTAACGAGTACTTGCATCTCTTCAAATTTACTTAATTGATCAAGTAAGTAGCAAAA

B



GAATTCTGTTGCCCTATTGCGCTGTAAGTTGCTAGTTGCAAGTTGCAAGTTGCACCTTTCTGCAGTTGATTTCTCCTCATCCACCTATGCAGTCAGGTGA
EcoRI HinPI

GAGGGAGTGAGTGCGAGTGGAGTGCTGAGGTGTGTCAAGCGAATTATTTATAAGGCCTAGAAGAAGGCAGCTCGCACGCGAATAATCAAGACTCAGCACC

AATTTTTAGTTTATGGTCTAGTTCTTTATAGGTTTTGTACTTCTTTTTTTTGCGTTGGCTATTTTGCGATTGAATTC
EcoRI

Recently, the 6778 bp long nucleotide sequence at the *D. mulleri Adh* locus has been reported (Fisher & Maniatis, 1985). The sequence analysis revealed that this region contained the *Adh-1* and *Adh-2* genes and an *Adh* pseudogene (*Adh*-ψ), all of which possessed single transcriptional initiation sites.

Sequence homology was also examined between the 1100 bp upstream region of the *D. melanogaster Adh* gene and the upstream region of the *D. mulleri Adh-1*, *Adh-2* or *Adh*-ψ genes. Little sequence homology was found. The upstream region of the pseudogene *Adh*-ψ, however, appears to have some homologous sequences around the -500 region and the TATA-box (Fig. 3.3.2). In spite of the homology, these homologous sequences are not likely to be functional elements for transcription because the *Adh*-ψ pseudogene does not appear to produce any transcripts (Fisher & Maniatis, 1985).

The transcriptional pattern of the *D. melanogaster Adh* adult and larval units during development is very similar to that of the *Adh-2* and *Adh-1* genes of *D. mulleri*. However, the lack of homology between the upstream regions of the *Adh* genes also suggests that the construction of the *D. melanogaster* dual promoter *Adh* gene is not directly derived from sequences utilized in the alternative expression of the *D. mulleri Adh* genes.

Fig. 3.3.2 Nucleotide sequence comparison of the 5′upstream region of the *D. melanogaster Adh* gene and the *D. mulleri Adh*-ψ gene.

The 1,100 bp upstream of the *D. melanogaster Adh* gene (the upper sequence) and the 550 bp upstream of the *D. mulleri Adh*-ψ gene (the lower sequence) were aligned to show sequence homology using the NUC-ALN program (Wilbur & Lipman, 1983). A part of the alignment is shown, which exhibits a high sequence homology. Vertical lines and dots indicate matching bases.

```
                                                                                -600
AGGGGGCGTGTTGTGCTTGCTTTCCAACTTTTCTAGATTGATTCTACGCTGCCTCCAGCAGCCACCCCTCCCATCACCATCCCCATCACCATCCAGTCCC

------------------------------------------------------------------------------------------------------

                                                                              -500
GTTGGCTCCCAGTCACAGTATTACACGTATGCAAATTAAGCCGAAGTTCAATTGCGACCGCAGCAACAACACGATCTTTCTACACTTCTCC--TTGCTAT
                                                  ::::::: :: ::: :: :  :   :::: ::: :  :  :::::
-------------------------------------------------TGCGACCACAACAAAAATAACAAAGTTCTTGACTACGTCGATTGCT--

                                                                            -400
GCTTGACATTCACAAGGTCAAAGCTCTTAATATTCTGGCTCGTGGCCCTACACTGTAAGAAATTACTATAGAAATAACGGTACACGGAATAAGATATTTT
  ::::::::::
--TTGACATTCAGTGAAA------------------------------------------------------------------------------------

                                                                            -300
TTTTAGTCCATATGCTTTTAACAAATGTGTTTTGAGTTTATGTTATATTATTGTTAGAAAACAGGTGTTTTTTTTTAAATCGGTTAAAAAATTACTACGA
                                                                          :::::::     :: :
------------------------------------------------------------------------GGTTAAATCTCCTCTCCAT

                                                                            -200
GAGAAAAATACAAATTTTGTAAATAAGATTGACTCTTTTTCGATTTTGGAATATTTTCATTCATTTTATGTTTTTACGTTTTCACTTATTTGTTTCTCAG
:  :      :      :           :::: : ::    :                                 :::::  : :  : :  : :::
GCTAGCTGCGAACTGCAAATGCGCTCCTCCCTCTCTATCTCTCACTC-------------------------------TTTTCCATCAACTATCGCACAG

          .         .         .         .         .         .         .         .         .
TGCAC----TTTCTGGTGTTCCATTTTCTATTGGGCTCTTTACCCCGCATTTGTTTGCAGATCACTTGCTTGCGCATTTTTATTGCATTTTACATATTAC
::: :    ::::         :: :    :  :  : :      :  : ::: :    :    ::::          ::::   : :  ::::::
TGCTCGATGTTTCCATATACATATGTGTATGTATGTGCATGCGTATGTGTGTGTATCGCTGTGTACTGCTGCATATCGCCTATTTGCTCTCCCATATTTA

                  .         .         .         .         .         .                   .
A------------CATTATTTGAACGCCGCTGCTGCTGCATCCGTCGACGTCGACTGCACTCGCCCCCACGAGA---------GAACAG------TATTT
           ::::: ::  ::                  ::::: :::::  :   :::: ::::::          ::::      :::::
CAGCAGACGATTTCATTACTTAGACA------------------TCGACCTCGACATCGACCGCCTCCACGACTTAGAGACCTGAACTATGTAACTATTT

      .                                                                                          .
AAGGAGCTGCGAAGG-------------------------------------------------------------------------------TCC
:::: ::::: :::                                                                                  :::
AAGGCGCTGCCAAGTTGGCGTCAAAGTCGACGTCGATGTCGACGTCAGCGTTGACTTCGCTGCCGGGCTATTTGATAAGTCAGCTTAAAAATTTTCCTCC

          X
AAGTCACCGATTATT
::   :::::
AAT--ACCGACAGCCTCGACAGCACGATTACCAATAACAGAGAAATAGTTCGACTTTTAGTAATCCTCAACATGCACTTCACTTTTTACTCTGGACAGGG

AAGCGGCG
       X
```

3.3.4 Distribution of thermodynamic stability of duplex DNA at the *D. melanogaster* and *D. mulleri Adh* loci

The thermodynamic stability of duplex DNA on the *D. melanogaster* and *D. mulleri Adh* loci was analysed by computing the free energy of base pairs in their nucleotide sequence. The values of free energy of ten nearest-neighbour interactions have been obtained from a thermodynamic data library (Breslauer et al., 1986; listed in Table 3.3.1). The value of the free energy was plotted against position in both *Adh* loci using a computer plotting program. As shown in Fig. 3.3.3, several striking features were found about the distribution of free energy: (1) The thermodynamic stability is relatively high in the coding regions of every *Adh* gene, while low in the 5′- and 3′-flanking regions or in the extragenic regions. (2) The stability is quite low in the introns and apparently distinguishable from that in the exons. (3) In the 3′-untranslated region, the stability is relatively low. (4) The stability in the presumptive coding region of the *Adh*-ψ pseudogene is low. However, the *Adh*-ψ has no region possessing such low stability as the introns in the real *Adh* genes. (5) The stability is high within approximately 250 bp of the 5′- ends of the *Adh* genes, while such a distribution of stability is not apparent upstream from the *Adh*-ψ. (6) The stability is relatively high from 600 to 800 bp upstream from the *D. melanogaster Adh* gene compared to that in other extragenic regions.

Table 3.3.1
Free energy of thermodynamic stability for nearest-neighbor interactions

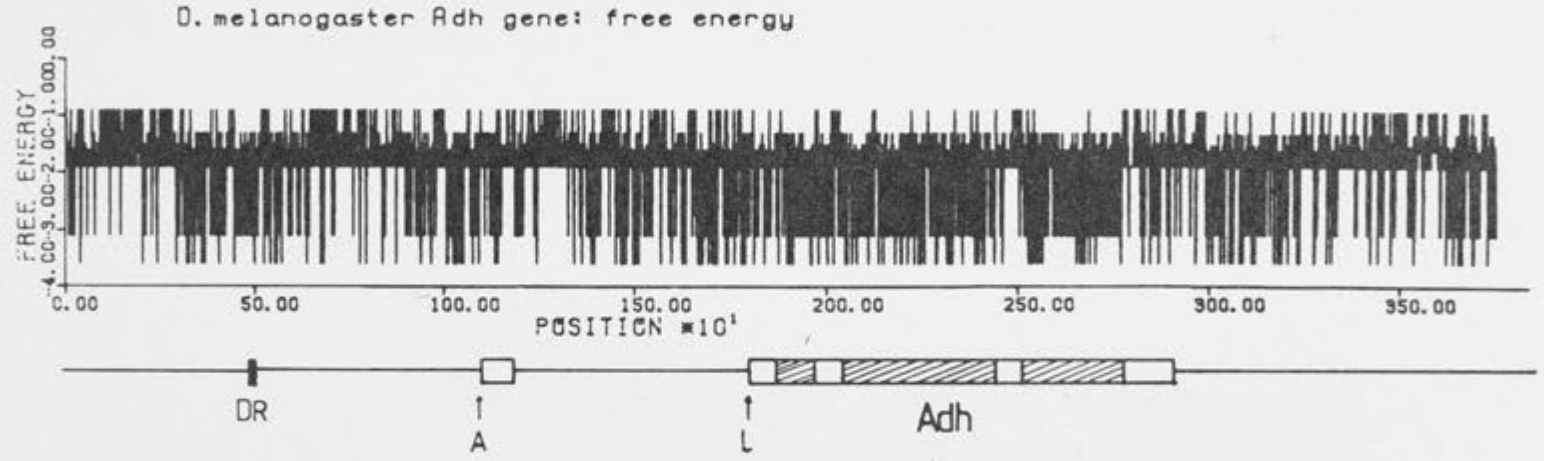
nearest-neighbor interaction	ΔG° [a]
AA/TT	1.9
AT/TA	1.5
TA/AT	0.9
CA/GT	1.9
GT/CA	1.3
CT/GA	1.6
GA/CT	1.6
CG/GC	3.6
GC/CG	3.1
GG/CC	3.1

[a] All values refer to the disruption of the interaction in an existing duplex at 1 M NaCl, 25°C, and pH 7. The unit for ΔG° is Kcal/mol of interaction. (This table is reproduced from Breslauer et al (1986).)

Fig. 3.3.3 **Thermodynamic stability of duplex DNA of the *D. melanogaster* and *mulleri* *Adh* loci.**

The free energies of thermodynamic stability of nearest-neighbour interactions are plotted against their positions at the *D. melanogaster* (A) and *mulleri* (B) *Adh* locus. The unit of free energy (ΔG°) is Kcal/mol. Open and hatched boxes represent the untraslated- and translated-regions of each *Adh* gene. A and L indicate the initiation sites of two transcripts of the *Adh* gene, respectively. DR indicates a 12-nucleotide direct repeat (see the text in detail).

A.


D. melanogaster Adh gene: free energy
FREE ENERGY
-4.00 -3.00 -2.00 -1.000.00
0.00 50.00 100.00 150.00 200.00 250.00 300.00 350.00
POSITION ×10¹
DR
A
L
Adh

B.

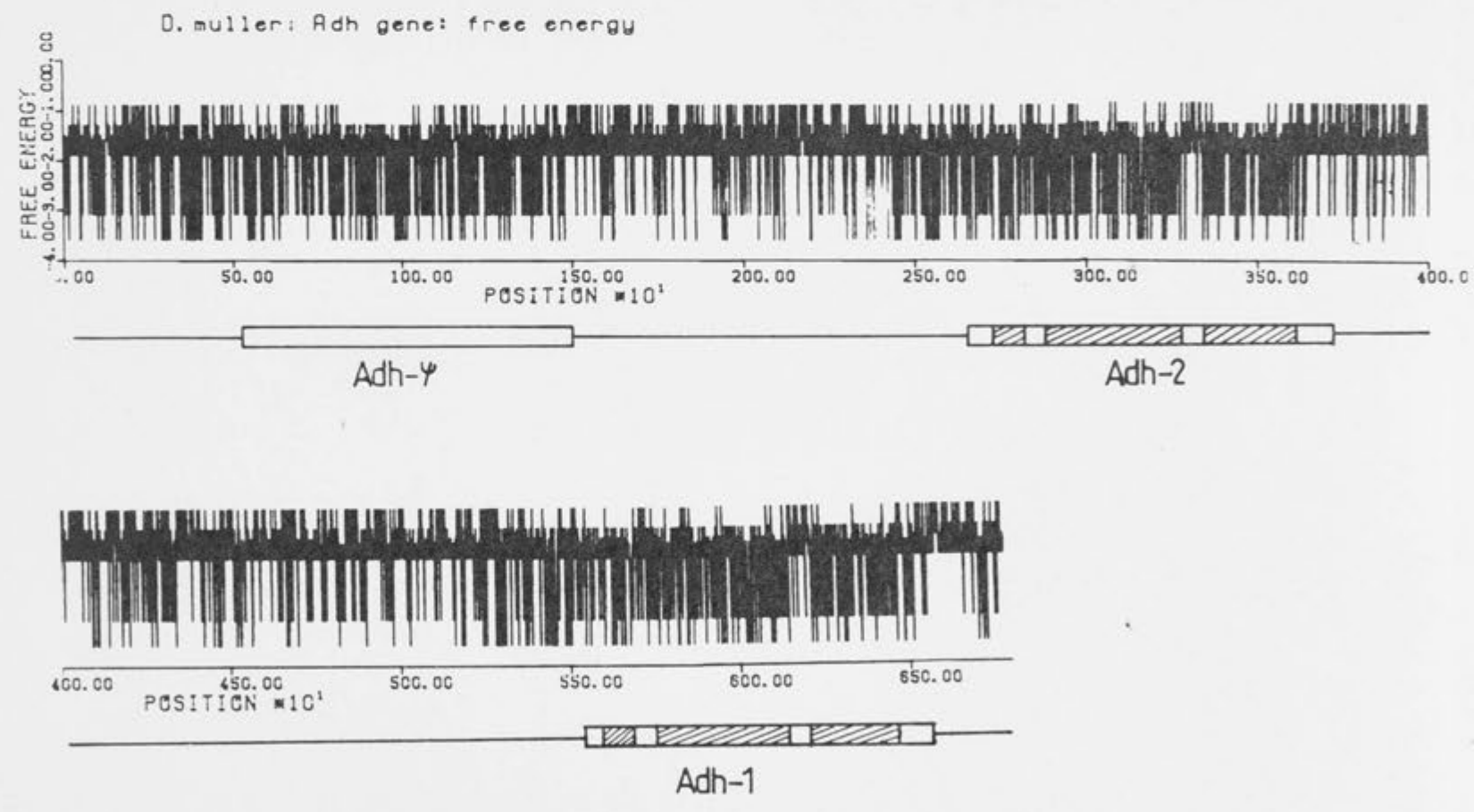

D. muller: Adh gene: free energy
FREE ENERGY
-4.00 -3.00 -2.00 -1.000.00
.00 50.00 100.00 150.00 200.00 250.00 300.00 350.00 400.0
POSITION ×10¹
Adh-Ψ
Adh-2
400.00 450.00 500.00 550.00 600.00 650.00
POSITION ×10¹
Adh-1

Fig. 3.3.4 Presumptive loops at 12 nucleotide direct repeat in the 5′upstream region of the *Adh* gene.

The nucleotide sequence surrounding the 12-nucleotide direct repeat is shown at the top. Numbers indicate the positions of bases relative to the adult capping sites, +1. The two possible loop structures formed by sequential slippage are shown below.

A

```
-660          -640          -620          -600          -580
TCTAGATTGATTCTACGCTGCCTCCAGCAGCCACCCCTCCCATCACCATCCCCATCACCATCCAGTCCCGTTGGCTCCCAGTCACAGTATTACACGTA
AGATCTAACTAAGATGCGACGGAGGTCGTCGGTGGGGAGGGTAGTGGTCGGGGTAGTGGTAGGTCAGGGCAACCGAGGGTCAGTGTCATAATGTGCAT
```

B

```
-660          -640          -600          -580
TCTAGATTGCTTCTACGCTGCCTCCAGCAGCCACCCCTC CCATCCCCATCACCATC-CAGTCCCGTTGGCTCCCAGTCACAGTATTACACGTA
AGATCTAACTAAGATGCGACGGAGGTCGTCGGTGGGGAG-GGTAGTGGTAGGGGTAG GTCAGGGCAACCGAGGGTCAGTGTCATAATGTGCAT
```

C

```
-660          -640          -600          -580
TCTAGATTGCTTCTACGCTGCCTCCAGCAGCCACCCCT CCCATCACCATCC-AGTCCCGTTGGCTCCCAGTCACAGTATTACACGTA
AGATCTAACTAAGATGCGACGGAGGTCGTCGGTGGGGA-GGGTAGTGGTAGG TCAGGGCAACCGAGGGTCAGTGTCATAATGTGCAT
```

It is noticeable that the distribution pattern of the thermodynamic stability is very similar among the *Adh* genes of the *D. melanogaster* and *D. mulleri*. In contrast, the pattern of the *Adh*-ψ pseudogene is slightly different.

Although the nucleotide sequence in the second intron is not well conserved among the *Adh* genes (Fisher & Maniatis, 1985), all the *Adh* genes similarily have a low stability in the second intron. These phenomena suggest that the distribution of thermodynamic stability will be conserved indirectly through the nucleotide sequence. Further, the distribution may have an important function in gene expression. A pseudogene will gradually lose such a distribution after its inactivation.

3.3.5 Fine mapping of altered DNA structures

In order to determine the accurate location of nuclease hypersensitive sites under torsional stress, the following experiments were performed by using nicked circular, but not linealised DNA molecules. A supercoiled psAc2 DNA was partially digested by S1 nuclease to produce about 70% nicked circle. The nicked psAc2 DNA was then cut with Xba I and labelled at the 3'- ends of the Xba I sites by Klenow polymerase reaction, followed by re-digestion by a second enzyme, Hind III. The Xba I - Hind III fragments were separated on agarose gel, then isolated. These fragments were electrophoresed on sequencing gels with a DNA marker to detect nicked sites in the fragments. From the analysis of the 1.15 kb Xba I - Hind III fragment,

several S1 hypersensitive sites were detected in the region of -420 to -560 (counting the nucleotides from the adult capping site (+1)) (Fig. 3.3.5).

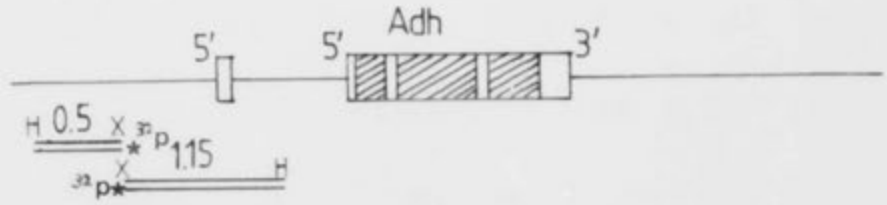
It should be mentioned here that this Xba I - Hind III fragment contains long dT-clusters and long dA-clusters, and further that a 12 nucleotide-long sequence (CCCATCACCATC) is repeated in tandem. Interestingly, 12 and 6 nucleotides in this repeated sequence are considered to be capable of adopting a loop out structure by sliding at two steps. These possible loops are illustrated in Fig. 3.3.4. In spite of the fact that such dT- or dA- rich sequences and direct repeat sequences are suspected to be major targets of S1 nuclease attack in a supercoiled DNA molecule, neither the dT- , dA- cluster nor the direct repeat in the Xba I - Hind III fragment were susceptible to nicking by S1 nuclease. Further, nucleotide sequences at the S1 nicking sites possess no obviously unusual features, yet these sites were specifically and reproducibly nicked by S1 nuclease when the DNA molecule was under torsional stress.

The S1 hypersensitive sites at this specific region were further analysed in detail using psAc1- topoisomers. The topoisomers were partially digested by S1 nuclease or T4 endonuclease VII, and the resulting nicked circular DNA was treated as described previously. The results are shown in Fig. 3.3.6. It indicats that S1 nuclease-susceptibilities at particular sites are complicated functions of the superhelical density of the DNA molecule.

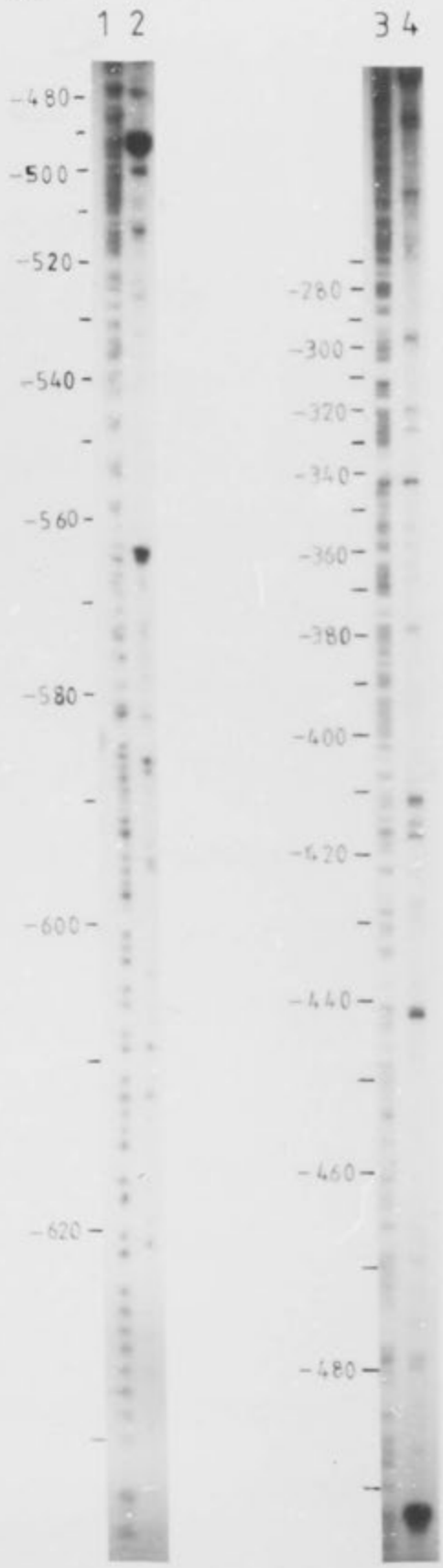
Fig. 3.3.5 Fine mapping of S1 nuclease cleavage sites in the 5′upstream region of the *Adh* gene.

The supercoiled *Adh* DNA, psAc2, was partially digested with S1 nuclease as described in Chapter 3.3.2, and cleaved by XbaI, and labeled with ^{32}P at the 3′ end of the XbaI site. The resulting 1.1 kb XbaI-HindIII fragment was isolated and analyzed on sequencing gel (B). The 0.5 kb XbaI-HindIII fragment was also isolated and analyzed (C). These fragments are schematically shown (A). The asterisks represent the ^{32}P-labeled ends of respective fragments. Lane 2; short run, lane 4; long run. The marker is Maxam-Gilbert A>C reaction on the XbaI-HindIII fragment (lanes 1 and 3). Numbers indicate the positions of bases. The cleavage sites are indicated on the sequence in Fig. 3.3.1.

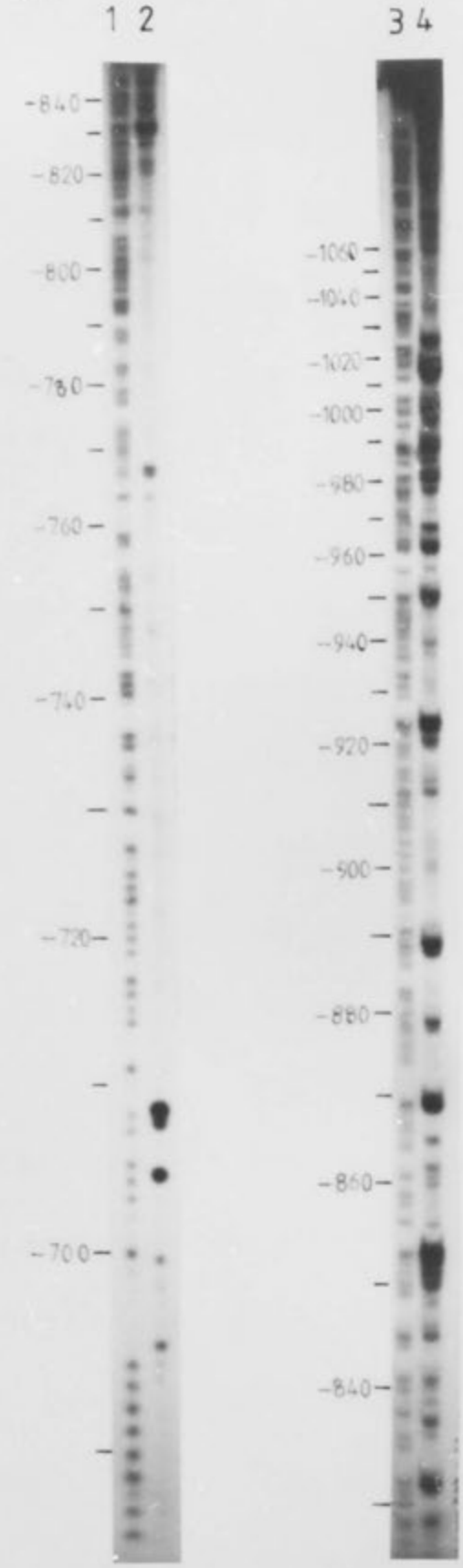
A.


5'
5'
Adh
3'
H 0.5 X
1.15
H
X

B.


1 2
3 4
-480
-500
-520
-540
-560
-580
-600
-620
-260
-300
-320
-340
-360
-380
-400
-420
-440
-460
-480

C.

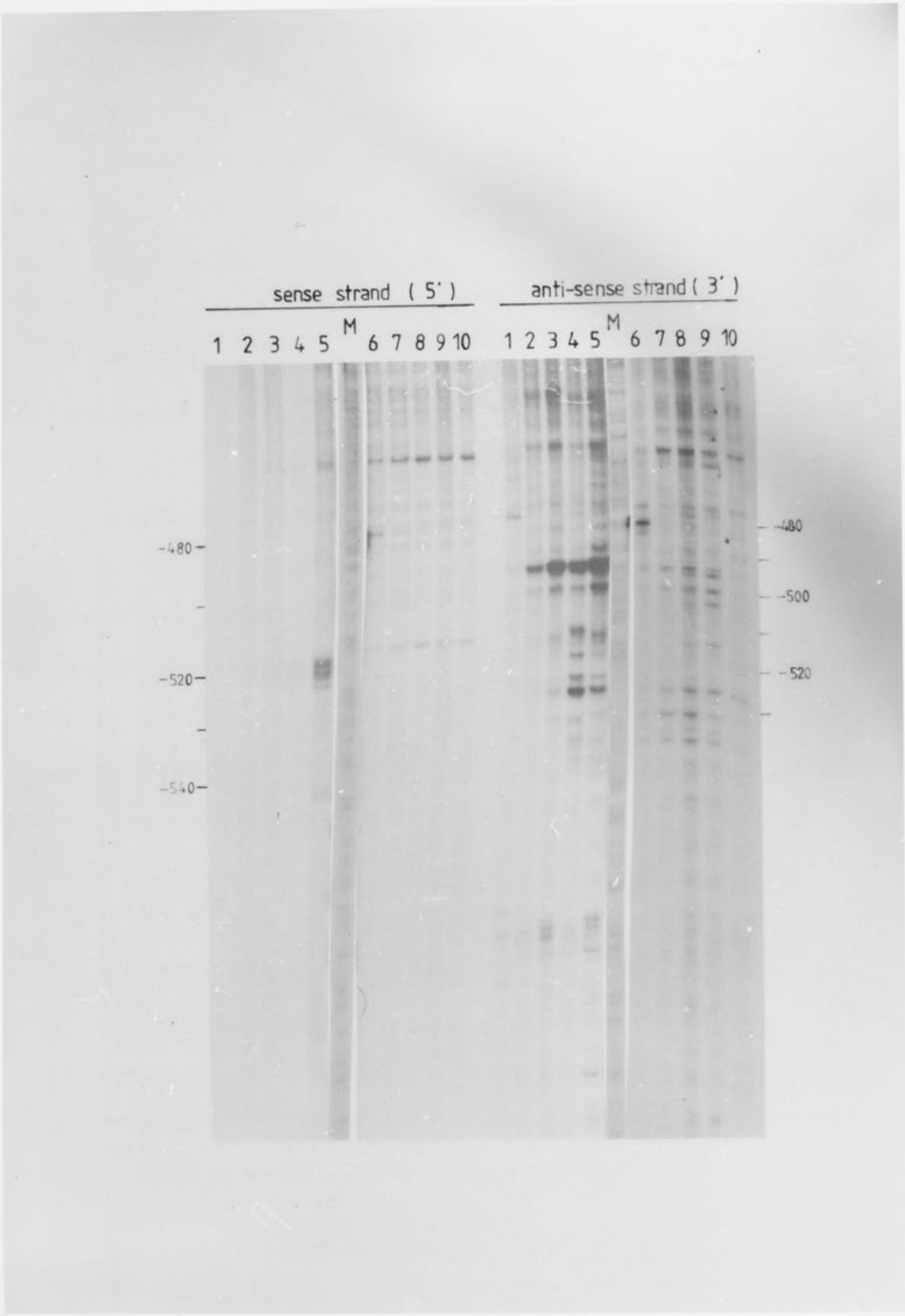

1 2
3 4
-840
-820
-800
-780
-760
-740
-720
-700
-1060
-1040
-1020
-1000
-980
-960
-940
-920
-900
-880
-860
-840

The susceptibilities at most of the sites were enhanced with an increase of superhelical density. At the -524 site on the anti-sense strand (the opposite strand to the coding strand of the *Adh* gene), however, the susceptibility was relatively high when the density ($-\sigma$) was 0.049. At $-\sigma$=0.049, further two S1-cleavage sites were newly detected at positions -516 and -511, while the site at position -512 was undetectable. Thus, the mode of susceptibility at the -520 site was different from that at other sites. The pattern of such superhelical density-dependent alteration of S1 nuclease susceptibility appears to be site-specific. Furthermore, the susceptible sites on both strands were not symmetric. On the sense strand (the same strand to the coding strand of the *Adh* gene), the S1 hypersensitive sites were located at positions -517 and -543. The susceptibility at these sites on the sense strand became higher according to the increase of superhelical density, and no site was newly detected depending on the density. Such an asymmetric susceptibility between strands excludes the possibility that the DNA duplex at this specific site is simply separated into a single stranded "bubble". This is further supported by the experiments using endonuclease VII as a specific tool.

Endonuclease VII can recognise a Holliday structure or a strand-branched structure (including a cruciform structure) of a DNA molecule (Mizuuchi et al., 1982). Supercoil-dependent sites susceptible to endonuclease VII were detected at positions -477 and -512. The site of -477 was detectable on both strands at high superhelical density

Fig. 3.3.6 Fine mapping of supercoil-dependent cleavage sites by S1 nuclease and T4 endonuclease VII downstream from the XbaI site in the 5'upstream region of the *Adh* gene.

The psAc1-topoisomers were partially digested with S1 nuclease and T4 endonuclease VII, respectively, and cleaved by XbaI. The XbaI sites are labeled with ^{32}P at either 3' or 5' ends. The 1.1 kb XbaI-HindIII fragment was analyzed as described in the legend to Fig. 3.3.5. The superhelical densities ($-\sigma$) of topoisomers are 0 (lanes 1 and 10), 0.025 (lanes 2 and 9), 0.037 (lanes 3 and 8), 0.049 (lanes 4 and 7) and 0.055 (physiological, lanes 5 and 6).


sense strand (5')
anti-sense strand (3')
1 2 3 4 5 M 6 7 8 9 10
1 2 3 4 5 M 6 7 8 9 10
-480-
-520-
-540-
-480
-500
-520

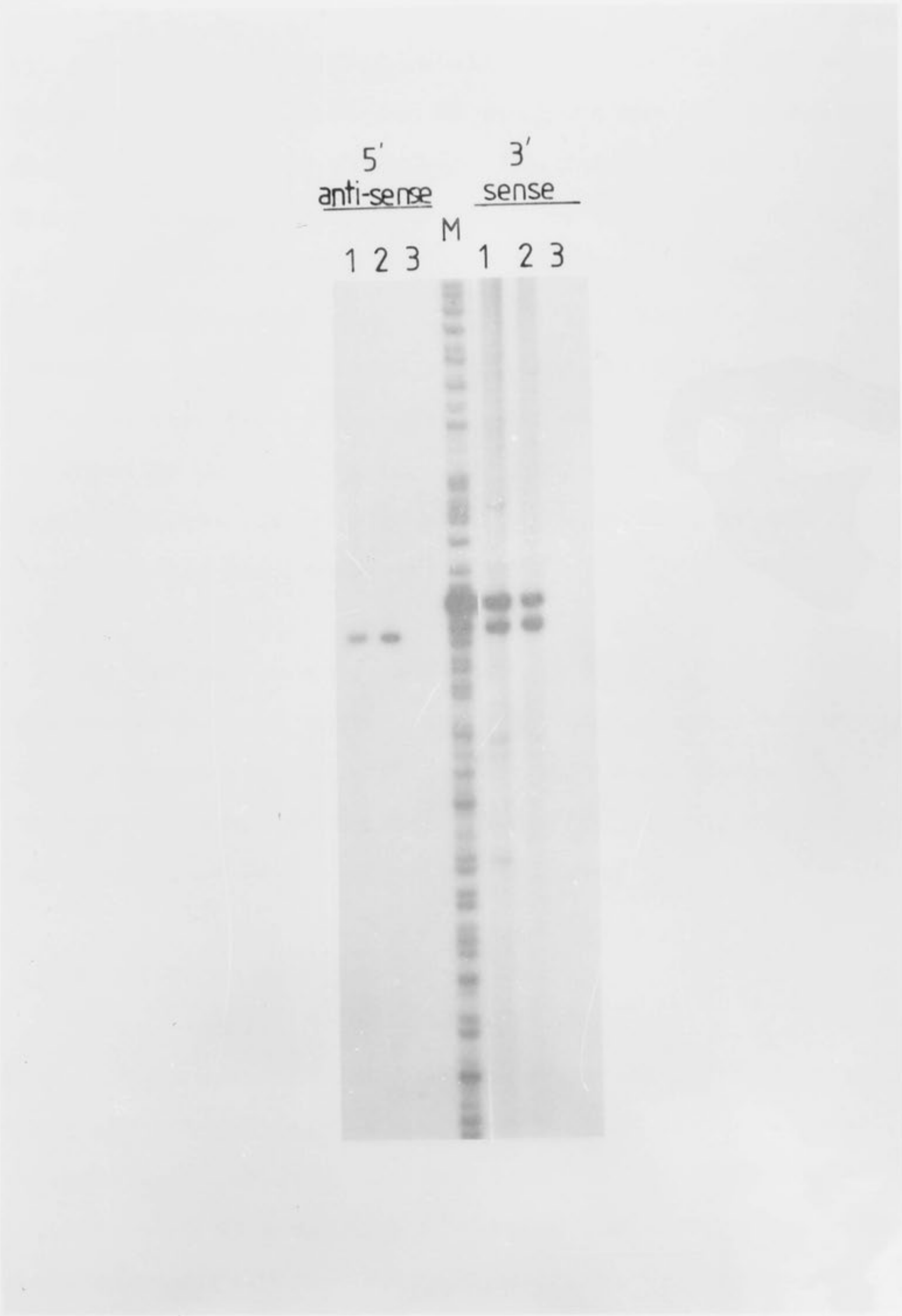
($-\sigma$ = 0.055), whereas the other cleavage site was detectable at position -512 only on the sense strand at lower density ($-\sigma$ = 0.025 ~ 0.037; lanes 8 and 9). The detection of site -512 together with the S1 sensitive site of -524 and -510, indicates that a peculiar structure is formed around the sites in slightly relaxed supercoiled DNA. However, the actual secondary structure around the sites remains ambiguous. In spite of the susceptibility to endonuclease VII no strand-branched structure was deduced from the nucleotide sequence.

In order to analyse S1 hypersensitive sites present in the region further upstream of the *Adh* gene, I used the 0.5 kb Hind III - Xba I fragment. Unexpectedly, a number of S1 hypersensitive sites were detected in this region (Fig. 3.3.7). Most of the detected sites were clustered in the region from -800 to -1020, while a major hypersensitive site was located at position -710, apart from the clustered sites. Neither of these sensitive sites possesses specific unusual features. However, it should be noticed that these sites are contained in a thermodynamically extremely unstable region of DNA duplex.

It seems likely that DNA duplex in this region is transiently melted to be single-stranded under the torsional stress of negative supercoiling. In such a melted region, S1 nuclease may recognise some preferential sites. Such a mechanism may allow hypersensitive sites to be clustered over nearly 200 bp.

Fig. 3.3.7 Fine mapping of cleavage sites by S1 nuclease and T4 endonuclease VII at 2.7 kb upstream of the *Adh* gene.

The supercoiled *Adh* DNA, psAF2, was partially digested with S1 nuclease (lane 2) and T4 endonuclease VII (lane 1). Alternatively, the DNA was not digested with any nuclease (lane 3). The DNA was further cleaved by EcoRI, and labeled with ^{32}P at the 3′ or 5′ ends of the EcoRI sites. The resulting 0.3 kb EcoRI-EcoRI fragment was isolated, further cleaved with HinPI and analyzed on sequencing gel.


5'
anti-sense
3'
sense
M
1 2 3
1 2 3

The S1 hypersensitive sites 2.7 kb upstream from the *Adh* gene was also mapped at nucleotide level. A supercoiled psAc2 DNA was partially digested to produce nicked circular DNA by S1 nuclease or endonuclease VII as described above. The nicked DNAs were then cut with Eco RI, followed by labelling at either 5′- or 3′- ends of the Eco RI sites. A 0.3 kb Eco RI fragment was isolated and redigested with HinP I enzyme. The Eco RI - HinP I fragment was analysed on a sequencing gel to detect nicked sites. The result is shown in Fig. 3.3.8. A major site was detectable on the anti-sense strand (the 5′- labelled strand) by S1 nuclease and endonuclease VII, while on the sense strand, two sites, which were only 3 bp apart, were detected by both nucleases. The positions of the hypersensitive sites were found to be nearly the same on the sense and anti-sense strands.

No obvious unusual sequences, e.g. polypurine·polypyrimidine, inverted repeat and alternating purine-pyrimidine, were formed around these sites. Further, it was noticed that T-clusters in this region were resistant to digestion by both nucleases.

3.3.6 Discussion

I determined the nucleotide sequences of the regions upstream the *Adh* gene, where specific DNA conformational changes were involved (see chapter 3.2). I also made a fine mapping of positions of altered DNA structures, which were specifically recognised under torsional stress by the

nucleases.

Firstly, the altered structure around position -550 exhibits a peculiar conformational property; it is detected in the presence of Mg(II) ion or on a slightly relaxed topoisomer by S1 nuclease, and it is also recognised by a single strand binding protein (SSB) only at a high concentration (see Chapter 3.2). An analysis of nucleotide sequence showed the presence of several dA- or dT- clusters and a 12 bp long direct repeat in this region, but none of them were responsible for the formation of altered DNA structures under torsional stress. Furthermore, neither Holliday nor strand-branched structures were revealed in this region from the nucleotide sequence in spite of a slight digestion by endonuclease VII. Strikingly, the altered fine structure in this region seems to be specifically changeable depending upon the superhelical density. Therefore, these observations indicate that the sequence involved in the altered DNA structure at this site should be unique and that the mechanism underlying the formation of the altered structure is not solely due to a known nucleotide sequence.

Secondly, the following hypersensitive sites, i.e. 2.7 kb and 0.75 kb upstream from the adult capping sites, seem to represent the demarcation points of the intrinsic territories of the *Adh* gene (see Chapters 3.1 and 3.2). Their molecular structures formed under torsional stress appear to be different from each other. No common nucleotide sequences were found. The 2.7 kb upstream

hypersensitive site is located in the 0.3 kb Eco RI fragment. The dT-clusters present in the region are not susceptible to either nucleases, again.

The S1 hypersensitive sites around position -710 and a cluster of sensitive sites in the region -800 to -1020 were revealed near the major hypersensitive site, which was detected on an agarose gel by indirect end-labelling method (see Chapter 3.2). The altered DNA structure at this site was strongly detected under high torsional stress, and in the presence of Mg(II) ion by S1 nuclease, and in a wide range of pH by various nucleases. The results obtained in the fine mapping experiments suggest that the high susceptibility of this specific region in various conditions is due mainly to the presence of a cluster of S1 hypersensitive sites, not to a single highly sensitive site. Probably, at least one altered DNA structure can be formed at one of the clustered sites under the correct conditions. Furthermore, the enhancement of the susceptibility of this site by Mg(II) ion and torsional stress may be determined by cooperative formation of altered structures at the clustered sites.

Moreover, this specific site exhibits another striking feature in that the S1 nuclease hypersensitive region at this site becomes wider in an unidirectional manner according to the concentration of SSB in the reaction (see Chapter 3.2). I previously indicated that unidirectionality is due to the unidirectional binding of SSB. Now the cluster of S1 hypersensitive structures can

nicely account for the unidirectionality. Namely, a SSB molecule first would bind a single stranded part of the altered structure formed at position -710 under torsional stress to stabilise a denaturation bubble. This first binding would facilitate the subsequent, consecutive binding of SSB molecules to the neighbouring altered structure in a cooperative manner (Kowalczykowski et al., 1981). Thus, the denaturation bubble would grow unidirectionally from the original site upstream because of the presence of a cluster of altered DNA structures in this direction.

I would like to dicuss various models proposed so far in relation to the formation of S1 nuclease hypersensitive structures detected in the *Adh* gene region. Firstly, calculations of helical twist, base pair roll and main chain torsional angle have shown a deformation of DNA helix. The variant DNA form is accounted for by a stereo-chemical clash of consecutive purine bases on opposite strands, thereby opening one groove and compressing the other (Dickerson & Drew, 1981; Calladine, 1982). Such variant DNA was suggested to be related to functional signals, e.g. TATA- and CAAT- boxes, and DNase I hypersensitive structures of chromatin (Nussinov et al., 1984). However, I did not find any particular features of such variant DNA around the S1 hypersensitive sites in the *Adh* gene. Furthermore, I analysed many other sequences of S1 hypersensitive sites. However, no common features of variant DNA were obvious in the sequences. Therefore, the variant DNA model does not fit to the altered DNA

structures detected in the *Adh* locus.

Secondly, poly d(A).poly d(T) has been proposed as a non-B form (heteronomous) structure, where the two strands have significantly different conformation (Arnott et al., 1983). It is expected that the tendency to form the heteronomous structure will increase under torsional stress. However, poly d(A) or poly d(T) sequence in the *Adh* gene region was not recognised by S1 nuclease, though such heteronomous structure may form there. The possibility is not excluded that the heteronomous structures near the S1 hypersensitivity may indirectly affect the neighbouring sequences.

Thirdly, the potential in polypyrimidine-strands for Hoogsteen base pairing (non-Watson Crick base pairing) has been suggested to be associated with S1 nuclease sensitivity to produce an asymmetric cutting pattern (Lee et al., 1979; Pulleyblank et al., 1985). However, this is unlikely for the altered DNA structures detected in the *Adh* gene region, since most of them are similarily recognised at pH 8 by Bal 31 nuclease (Chapter 3.2), the protonation will not be induced at pH 8.

Finally, it is well known that cruciform structures at inverted repeat sequences and Z-form structures at alternating purine-pyrimidine sequences are induced under torsional stress to be recognised by S1 nuclease. The reduction in twist resulting from the formation of Z-form or cruciform structures decreases the magnitude of the

writhe component. The decrease will retard the mobility on electrophoresis (Wang et al., 1983). I analysed topoisomers of the psAc2 DNA or several small plasmid DNAs containing the S1 sensitive segments on a two-dimensional gel. The analysis, however, did not show any retardation in their mobilities (data not shown). This observation suggests that the S1 hypersensitive sites do not contain any alteration of the DNA duplex-handedness nor the duplex of an intrastrand by which the writhe number is decreased.

All of the neighbouring hexanucleotides of major S1 nuclease hypersensitive sites are listed in Table 3.3.2. No obviously common sequence is deduced from the list. However, either d(TT) or d(AA) is contained around most of the sensitive sites. It has been reported that the wedge angle between the base pairs in d(AA).d(TT) is larger than in other dinucleotide, and this large wedge angle is concerned with the curvature or bending of DNA duplex (Ulanovsky et al., 1986). However, not all such dinucleotides seemed to be sensitive to S1 nuclease, because poly d(A) in the *Adh* gene region was resistant to S1 nuclease attack.

Nucleotides flanking the dinucleotide probably affect its wedge angle. Assumably, the structure of such dinucleotide may be further deformed in by torsional stress and the deformed structure recognised by S1 nuclease. It seems plausible that in a thermodynamically unstable region, the dinucleotides may be easily deformed. The clustered S1 hypersensitive sites in positions -800 to

Table 3.3.2
List of hexanucleotides surrounding S1 nuclease cleavage sites detected at the *Adh* locus

fragment for detection[a]	label[b]	hexanucleotide[c]	
HinPI-EcoRI*	3′ (sense strand)	T↓AAGGC	GGCCTA↓
	5′ (anti-sense strand)	CTAGGC↓	
HindIII-XbaI*	3′ (sense strand)	-1027 ATATTT -1022	-1016 AAATTG -1011
		-1003 GTCAAT -998	-992 TTTAAG -987
		-984 ATTAAG -979	-966 TACAAT -961
		-953 CTTACC -948	-927 TTTGAA -922
		-892 TACGAA -887	-872 TTTAAA -867
		-856 ATTGCC -851	-845 ATTAAA -840
		-711 GCTAAT -706	-707 ATGAGC -702
XbaI*-HindIII	5′ (sense strand)	-547 TCAATT -542	-521 ATCTTT -516
	3′ (anti-sense strand)	-561 TTGCAT -566	-521 TCGTGT -526
		-490 GTCAAG -495	-418 CGTTAT -423

[a] (*); This mark represents the ^{32}P-labeled site.
[b] (3′); The 3′ end of site was labeled with ^{32}P by Klenow fragment reaction.
[c] (↓); Arrows represent major S1 cleavage sites. Short arrows represent minor S1 cleavage sites near a major site.

(5′); the 5′ end of site was labeled with ^{32}P by polynucleotide kinase

-1020 are indeed located in an extremely unstable region, but other sites are not. It is still unsure what actual secondary structures are induced by torsional stress.

Nevertheless, all sensitive sites in the *Adh* gene region seem to have their own conformational features. Torsional stress can induce various different types of secondary structure at specific sites. Some that different types of secondary structures may be involved in specific functions for the formation of higher order chromatin structure. It has been demonstrated that the locations of some S1 sensitive sites correspond to the demarcation points of intrinsic territories. Therefore, it is also possible that the location of these secondary structures may significantly contribute to the formation of such chromatin structure by a common feature.

I have demonstrated that the distribution of thermodynamic stability along the *Adh* genes of *D. melanogaster* and *D. mulleri* is very similar. The distribution seems to be conserved through the rearrangement of the *Adh* gene during evolution. By contrast, the nucleotide sequence of the 5'- flanking region is not conserved among the *Adh* genes. This observation suggests that a primary sequence in itself is not so important for gene functions other than protein coding but that the whole organisation of nucleotide sequences secondary configuration is more important.

Of course, such a secondary configuration is

indirectly reflected by a primary sequence. I have not analysed altered DNA structures in the *D. mulleri Adh* loci yet. It seems likely to me that similar altered DNA structures to that in the *D. melanogaster* will be revealed at or near the same positions.

3.4 *In vitro* transcription of superhelical topoisomers

3.4.1 Introduction

Evidence has been well documented that gene transcription in prokaryotes and eukaryotes is regulated by trans-acting mechanisms, such as promoter-activator recognition. Regulator proteins can interact specifically with operators that often exist in the 5′- flanking region of a gene. In a few cases, e.g. 5S RNA gene and t-RNA genes, regulatory operators for transcription were observed in the intergenic region. On the other hand, regulation of gene expression may also operate by a cis-acting mechanism which involves the conformational change of the DNA in a gene. Recently, a unique regulation mechanism for transcription has been proposed (Hochschild & Ptashne, 1986). Proteins bound at widely separated sites contact each other to form an intervening DNA loop or bend, which allows the protein-protein interaction for transcription.

In this context, I would like to mention another mechanism, cis-acting through torsional stress. In *in vitro* systems of prokaryotes, a closed circular DNA has much higher transcriptional activity as template than a linear DNA (Wang, 1974). Further, similar observations in a eukaryotic system have recently been made by microinjection experiments. A circular template DNA microinjected into oocyte nuclei was much more transcriptionally active than a linear template (Probst et

al., 1979; Harland et al., 1983). These observations imply that the conformational state of template DNA plays an important role in efficient recognition by transcriptional factor(s) or RNA polymerase. In fact, RNA polymerase II purified from wheat germ binds to supercoiled Col E1 DNA more efficiently than to linear DNA (Lilley & Houghton, 1979). RNA polymerase II binds to a promoter site to form a transcriptional initiation complex prior to the elongation of transcripts. On the binding of RNA polymerase II, helical unwinding of the DNA at a promoter site occurred in conjunction with the formation of an open promoter (Gamper & Hearst, 1982, Beard et al., 1984). Furthermore, TFIIIA transcriptional factor binds at the intragenic site of the 5S RNA gene for active transcription. This factor also unwinds the DNA duplex at the binding site through the formation of a TFIIIA-DNA complex (Gottesfeld & Bloomer, 1982; Reynold & Gottesfeld, 1983; Hanas et al., 1984). Thus, these observations raise the possibility that not only circularity but also superhelicity of template DNA are required for efficient transcription. This possibility is also supported by some experiment using topoisomerase inhibitors (see Chapter 1.2).

In this experiment, I prepared an *in vitro* transcription system containing whole cell extract from the mouse sarcoma 180 (TG) cells (Manley et al., 1980). Using this system, I assayed the transcriptional activity of the *D. melanogaster Adh* gene contained in recombinant DNAs which possessed various superhelical densities. The results have shown that highly superhelical DNA has an

efficient transcriptional activity as a template.

3.4.2 Materials and methods

(a) Cell free extracts for *in vitro* transcription

Mouse sarcoma 180 (TG) cells were maintained by inoculation every week. The cells were collected 5 days after inoculation unless specially mentioned. A cell free extract for *in vitro* transcription was obtained by a slightly modified procedure from Manley's methods (Manley et al., 1980). The sarcoma 180 (TG) cells were washed with phosphate buffered saline several times and suspended with two packed cell volumes of hypo-buffer [10 mM Hepes (pH 7.9), 10 mM KCl, 1.5 mM $MgCl_2$ and 0.5 mM dithiothreitol]. The suspension was left on ice for 10 min to swell the cells and homogenized in a Dounce homogenizer by 7-10 strokes. One tenth volumes of 10 x S100 buffer [0.3 M Hepes (pH 7.9), 1.4 M KCl, 0.03 M $MgCl_2$] was added to the homogenate, then frozen and thawed four or five times.

The mixture was centrifuged for 60 min at 45,000 rpm in a Beckmann 50 Ti rotor (100 000 x g). The supernatent was dialyzed against dialysis buffer [20 mM Hepes (pH 7.9), 20% (v/v) glycerol, 100 mM KCl, 0.2 mM EDTA and 0.5 mM dithiothreitol] for four hours and stored at -80°C until use.

From chicken bone marrow cells, a cell free extract

was prepared as above. The bone marrow cells were collected from three week old chickens and washed with phosphate buffered saline several times to remove erthrocytes and fat.

(b) *In vitro* transcription of linear DNA

A 592 bp Hinf I - Hpa I fragment on the *Adh* gene locus was isolated from an agarose gel and purified through DEAE-Sephacel.

A standard *in vitro* transcription mixture (total volume 50 µl) contained 20 µl of cell free extract extract, 10 mM Hepes (pH 7.9), 0.33 mM dithiothreitol, 0.67 mM each ATP, GTP and CTP, 0.026 mM UTP, 10 µg of [α-^{32}P] UTP, 80 mM KCl, 2 mM Mg-acetate and 1 µg/ml of template DNA. The mixture was incubated at 30°C for 60 min. The reaction was terminated by addition of 200 µl of a stop solution [100 mM Na-acetate (pH 4.5), 0.6% SDS, 0.3 M NaCl, 50 µg/ml t-RNA] and 150 µl of distilled water. RNA was purified by two phenol extractions and ethanol precipitations. The RNA was glyoxalated in 1 M glyoxal, 0.1 M Hepes (pH 7.9), 50% (v/v) dimethylsulfoxide at 37°C for 2 hr, and electrophoresed on a 4% polyacrylamide gel. Size markers were provided by an end-labelled Hinf I and Eco RI digest of pBR332 DNA which was glyoxalated as above.

(c) *In vitro* transcription of circular DNA

One microgram of an circular *Adh* recombinant plasmid

DNA (psAc1) was transcribed at 30°C for 5 min in 25 µl of the same *in vitro* transcription mixture to that of the linear. The transcription reaction was terminated by an addition of 5 µl of a sarkosyl - EDTA solution to final concentrations 1.5% (w/v) and 20 mM respectively. This mixture was run on 1% (w/v) agarose gel in 1 x TAE buffer. After electrophoresis, the gel was dried on 3MM paper (Whatmann), and then exposed on an X-ray film.

To analyse the superhelical topology of the DNA, template DNA was extracted with PCIA once and chloroform once after the transcription reaction. The DNA was slowly run on 1% (w/v) agarose gel in 1 x TAE buffer in the absence of ethidium bromide (see Chapter 2). After electrophoresis, separated DNAs were detected under a shortwave UV light.

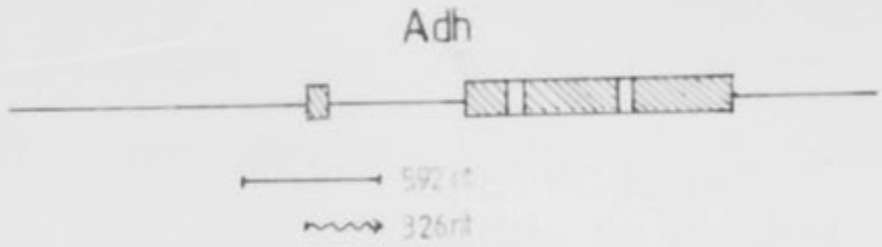
3.4.3 *In vitro* transcription of the *D. melanogaster Adh* gene in heterologous cell extract

The Hinf I -Hpa I DNA fragment (592 bp) containing the adult promoter and 5′- coding regions of the *Adh* gene (Fig. 3.4.1 A) was added as a template to the sarcoma 180 TG cell free extract, and incubated at 30°C. The incorporation of [α-^{32}P]-GTP into RNA transcripts was detected by polyacrylamide gel electrophoresis and autoradiography as described in Chapter 3.3.2. Two discrete sizes of transcript were detected, i.e., 592 base and 326 base long (Fig. 3.4.1 B). The longer transcript showed the same size as that of the Hinf I - Hpa I DNA fragment; therefore it

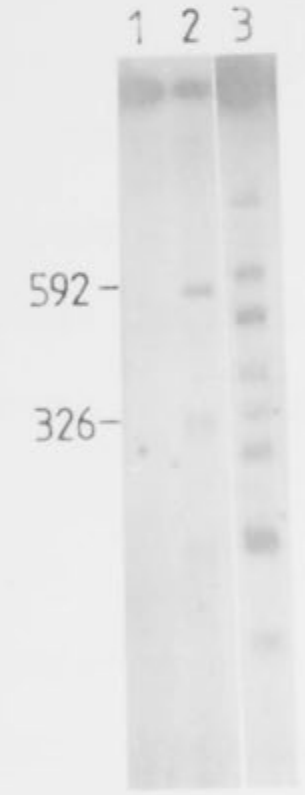
Fig. 3.4.1 *In vitro* transcription of the *Adh* gene on a DNA fragment in heterogeneous cell extract.

(A) The schematic map of the *Adh* gene. The HinfI-HpaI fragment (592 bp long) was isolated from psAc1 DNA. This fragment contains the adult promoter region. (B) Run-off *in vitro* transcription was carried out as described in Chapter 3.4.2. Transcripts were electrophoresed on 4% polyacrylamide gel. Reaction mixture contains 1 µg/ml of Adh template DNA (the HinfI-HpaI fragment) (lane 2) or no DNA (lane 1). Size marker was HinfI-EcoRI cut pBR322 labeled with ^{32}P, followed by treatment with glyoxal as for RNA sample (lane 3).

A.



B.



must be a non-specific transcript which was transcribed from end to end of the DNA fragment. The size of the shorter transcript, however, was identical to the distance from the adult capping site to the Hpa I site (Fig. 3.3.1 A). This observation indicated that the 362 bp transcript was produced by faithful transcription, initiated at the proper site and run off at the Hpa I site.

I prepared several batches of whole cell extracts from the sarcoma 180 (TG) cells grown in mice (see Chapter 3.3.2). It was noticed that the transcriptional activity differed in the cell free extracts from cells harvested at different periods after inoculation. At five days after inoculation, the cell free extract seemed to show the highest transcriptional activity. Since maximum cell growth was seen five days after innoculation, the transcriptional activity of the extract appeared to correlate with the growth activity of the cells. It, therefore, suggested that growing cells contain appropriate factor(s) necessary for efficient transcription. This suggestion is widely supported in other examples of cell free extracts from *in vitro* culture cells, e.g., HeLa cell (Manley et al., 1983).

A cell free extract prepared from chicken bone marrow cells was also used to examine the transcriptional activity of the *Adh* gene. In this extract, the Hinf I - HpaI DNA fragment of the *Adh* gene was used as a template. The *Adh* gene was transcribed faithfully from the proper initiation site (data not shown), though the transcriptional activity

was relatively lower than that in the sarcoma extract.

As described above, the *D. melanogaster Adh* gene is readily transcibed from the proper initiation site in various cell free extracts, which are prepared from phylogenetically heterogenous animals, i.e., mouse, chicken and human (Benyajati et al., 1982). This observation suggest that essential factors for the faithful transcription of the *Adh* gene are universally contained in many cells. Therefore, it seems that the factors and machineries essential for the transcription are common.

3.4.4 Supercoiled DNA is a more efficient template for transcription

To determine whether the superhelicity or circularity of a DNA template is required for active transcription, I added the circular form of psAc1 DNA containing the *Adh* gene to the cell free extract from sarcoma 180 (TG) cells. The superhelicity of the DNA was first analysed on an agarose gel at intervals while the supercoiled DNA was incubated in the *in vitro* transcriptional system at 30°C. The results showed that the supercoiled template DNA was rapidly relaxed within a minute, when incubated in the extract. However, the template DNA, once relaxed, was again found to be highly supercoiled, in the negative sense, after a longer incubation (30 min) (data not shown). It was deduced that on incubation of the template DNA in the *in vitro* system, the supercoiling of the DNA molecule was relaxed by the nick and turn reaction of topoisomerase

I present in the cell free extract, then assembled into chromatin structures in an association with histone and other proteins.

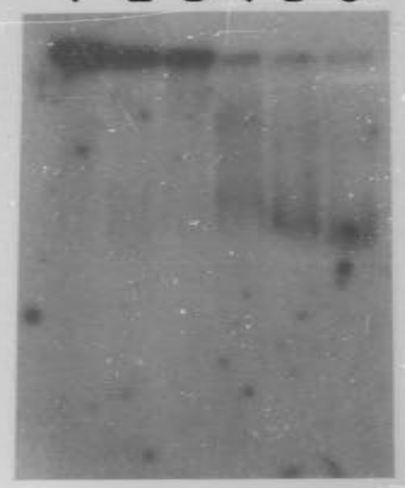
In further experiments, the incubation of the circular DNA was carried out for a limited period, i.e., five minutes, to minimize the transitional effects of superhelical conformation on the transcriptional activity of the template DNA. Nascent RNA being transcribed is known to form a stable ternary complex with template DNA and RNA polymerase (Losiok & Chamberlin, 1976). It has been reported that sarkosyl disintegrates chromatin structures, but ternary transcriptional complexes are resistant to attack by sarkosyl (Chelm & Geiduschek, 1979; Tolunay et al., 1984). After the supercoiled psAc1 DNA containing *Adh* was incubated for five minutes in the *in vitro* system, and treated with sarkosyl, then ternary complexes retaining nascent ^{32}P-labelled transcripts of the *Adh* gene were detected on agarose gels (Fig. 3.4.2 A). ^{32}P-labelled transcripts were associated with template DNA together with histone and other proteins and were retarded on the gel when not treated with sarkosyl. At 1.5% sarkosyl, some proteins were removed from the template DNA, and the nascent ^{32}P-transcripts, still tightly bound to the DNA, ran relatively faster on the gel. Further, when the transcript-DNA complex was treated by heating at 100°C for one minute, no ^{32}P-transcript bound to the template DNA was detected (data not shown). This result confirmed that the incorporation of [α-^{32}P] -GTP detected on the gel was due to nascent RNA-DNA complexes. This method therefore

Fig. 3.4.2 Detection of transcriptional ~~tanary~~ ternary complex on topoisomer DNA.

(A) The supercoiled Adh DNA, psAc1, was transcribed at 40 µg/ml at 30°C for 5 min in cell free extract as described in Chapter 3.4.2. The reaction was terminated by addition of EDTA to final concentration of 20 mM EDTA. Sarkosyl was added to the reaction mixture to the concentrations of 0 (lane 1), 0.25 (lane 2), 0.5 (lane 3), 0.75 (lane 4), 1.0 (lane 5) and 1.5% (lane 6). These mixtures were electrophoresed on 1% agarose gel. (B) The psAc1-topoisomers used in this experiment were the same as those in Fig. 3.2.3 (lane 1 to 10). The topoisomers were transcribed at 20 µg/ml for 5 min in cell free extract. EDTA and Sarkosyl were added to the reaction mixtures to the concentrations 20 mM and 1.5%, respectively. Those mixtures were electrophoresed on 1% agarose gel.

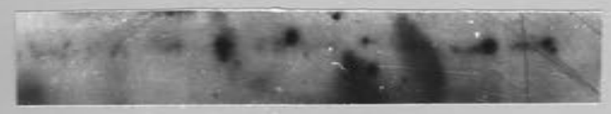
A.

1 2 3 4 5 6



B.

1 2 3 4 5 6 7 8 9 10



monitored the transcriptional activity of circular DNA.

A series of psAc1-topoisomers containing the *Adh* gene ($-\sigma$ = 0 ~ 0.055) were incubated in the *in vitro* transcriptional system for five minutes, and ^{32}P-labelled transcripts retained in the ternary complex were examined as above (Fig. 3.4.2 B). It was apparent that incorporation into transcripts of *Adh* increased as the superhelical density of the template DNA became higher. The results indicated that highly supercoiled template DNA was preferentially transcribed in the *in vitro* system, whereas the relaxed DNA was poorly transcribed.

3.4.5 Discussion

The above experiments demonstrated that the *D. melanogaster Adh* gene was faithfully transcribed in an *in vitro* transcriptional system which contains the cell free extract from the mouse sarcoma 180 (TG) cells or chicken bone marrow cells. It has also been reported that a cell free extract from human HeLa cells allows a faithful transcription of the *Adh* gene (Benyajati et al., 1982). Thus, the *Adh* gene is readily transcribed in cell free extracts from various sources, which are phylogenetically distant from *D. melanogaster*.

Further, many kinds of genes are well known to be transcribed in HeLa cell free extract (Manley et al., 1983). In *Xenopus* oocyte nucleus, additionally, transcription from many kinds of genes has been assayed by

microinjection experiments (Gurdon & Wickens, 1983). Recently, it has been widely reported that cloned genes in recombinant DNA are actively transcribed in cultured cells following DNA mediated transformation. The *D. melanogaster Adh* gene is also correctly expressed after transformation of *Drosophila* cells in culture (Benyajati & Dray, 1984).

The active expression of such cloned genes in recombinant DNA raises the possibility that the cloned genes are free from suppression mechanisms of transcription. Further, without such suppression mechanisms, only a limited number of transcriptional factors may be required for transcription of every eukaryotic gene in a similar manner to prokaryotic genes. Probably all genes are readily transcribed in a simple transcription system. In the transformed *Drosophila* culture cells, in fact, the endogenous *Adh* genes are not expressed while the exogenous *Adh* genes in cloned DNA are (Benyajati & Dray, 1984).

What regulates the expression of each gene in a tissue-specific and stage-specific manner if a limited number of factors are universally sufficient for active transcription of genes? It is known that enhancers, hormones and other tissue-specific proteins are involved in specific regulation in a positive manner. Little, however, is known about suppression mechanisms, though silencers and repressors have been observed. It is thought that the bulk of a eukaryotic chromosome is organised into a condensed structure, which does not permit recognition of promoters

or other regulatory elements by particular transcriptional factors. Perhaps this may exist in a basic state of suppressive regulation.

In the *in vitro* transcriptional system containing the cell free extract from sarcoma 180 (TG) cells, I have examined the transcriptional activities of superhelical topoisomers of the *Adh* gene. The result showed that the highly supercoiled template DNA was preferentially transcribed. Now we can say that supercoiling rather than circularity of template DNA was responsible for efficient transcription.

I previously demonstrated that the psAc1 DNA, containing the *Adh* gene, plastically changes its conformational structures, according to the superhelical density of the circular DNA molecule and its microenvironments (see Chapter 3.2). This observation suggests that altered DNA structures induced by torsional stress play an important role in efficient transcription.

In contrast, it has been reported that the efficiency of the SV 40 early gene transcription in an *in vitro* system increases on both circular and linear templates in the presence of the SV 40 enhancer sequence over hundreds of base pairs from its position (Sergeant et al., 1984). Neither circularity or torsional stress is necessary for enhancement of transcription by the enhancer sequence. The authors also suggested that their observations seem to reflect the binding of some elements of transcriptional

machinery to the enhancer sequences. This observation appears to be at variance with my results. However, regulation sequences other than enhancer sequences may require the conformational alteration of their DNA for the binding of such elements to these sequences.

It has also been suggested that the superhelical form of chromatin DNA is responsible for active transcription in SV 40 and *Xenopus* 5S RNA gene (Luchnik et al., 1982; Ryoji & Worcel, 1984). Further, Villeponteau et al. (1984) have proposed that altered structures around active genes are maintained by continuous DNA supercoiling. Thus, torsional stress is thought to function not only in the formation of altered DNA structures for binding of transcriptional factors, but also in the formation of active chromatin structures for transcription.

Under the conditions for *in vitro* transcription used here, the template DNA is considered to be essentially free of nucleosomes. The experiments on superhelicity indicated that the DNA in the cell free extract is assembled into chromatin over at least 30 minutes. This is consistent with others' observations (Sergeant et al., 1984 Glikin et al., 1984). However, I incubated the reaction mixture for only five minutes to measure the incorporation into transcripts on stable ternary complex. Therefore, the possibility can be obviously ruled out that altered DNA structures play a direct role in the formation of active chromatin structures for transcription in this system, because the incubation time is too short to form a chromatin structure.

It seems more likely that altered DNA structures are readily recognised by the factors required for efficient transcription.

3.5 Chromatin structures of the *D. melanogaster Adh* gene region during the switching of its dual promoters

3.5.1 Introduction

During development of eukaryotes, the transcriptional activities of many structural genes are strictly limited to specific tissues and also certain developmental stages. In prokaryotes, differential expression of genes commonly operates through trans-acting control mechanisms, which are readily reversible. Although such mechanisms also act upon expression of eukaryotic genes, it seems of significance that the developmentally controlled events may be mediated through very stable cis-acting chromosomal structures that can be stably propogated to progeny cells (reviewed by Weintraub, 1985). For many gene regions, these chromatin structures are currently under extensive scrutiny.

Coding regions and also flanking regions of genes are known to be sensitive to DNase I (Weintraub & Groudine, 1974). Such sensitive regions spread over quite a long distance, i.e. 10 to 100 kb (Lawson et al., 1982). Evidence has accumulated that DNase I hypersensitive sites often exist around active or potentially active structural genes (reviewed by Eissenberg et al., 1985). Domains of such active chromatins display several specific sites which are hypersensitive to DNase I, though not all DNase I hypersensitive sites are associated with active genes. It has been suggested that the acquisition of such DNase I hypersensitive sites around a gene is a pre-requisite for

transcriptional activation, but not a consequence of the process of transcription (Stalder et al., 1980 a & b; Groudine & Weintraub, 1982).

For example, some DNase I hypersensitive sites were detected, in stage-specific and hormone-dependent manner, at the 5′- flanking regions of the chicken vitellogenin gene (Burch & Weintraub, 1984) and lysozyme gene (Fritten et al., 1984). The DNase I hypersensitive sites in the 5′-flanking region of the chicken β-globin gene are detectable only in the erthrocytes and the erythroid precursor cells, not in non-erythroid cells (Groudine & Weintraub, 1981). Moreover, a series of DNase I hypersensitive sites in the 5′-flanking region of Sgs 4 *glue* protein gene is found specifically in the salivary gland (Shermoen & Beckendorf, 1982).

The structural features of DNase I hypersensitive sites may be defined by particular DNA sequences at or near the sites. It was found that a series of mutations upstream of the Sgs 4 gene, which decreased its expression, also diminished sensitivities to DNase I in this region (McGinnis et al., 1983). Furthermore, both the 72 bp and 21 bp repeats of the SV 40 genome appear to be responsible for generating DNase I hypersensitivity (Jongsta et al., 1984). These repeats possess alternating dG-dC sequences which can readily adopt a Z-form structure.

It was found in electron microscopic studies that the DNase I hypersensitive sites in the repeats are contained

in a nucleosome free region of the SV 40 minichromosome (Saragosti et al., 1980; Jakobovits et al., 1980). The 5′-flanking region of the chicken β-globin gene in globin-producing cells was readily digested by several nucleases, e.g., micrococcal nuclease, S1 nuclease, Msp I restriction enzyme and DNase I (McGhee et al., 1981). These observations suggest that the relevant conformational property of DNase I hypersensitive chromatin leads to freedom from nucleosomes.

Recent extensive investigations have been carried out on the nature of proteins which may be involved in the creation of DNase I hypersensitive sites. Such proteins appear to be tissue-specific and site-specific. For example, DNase I hypersensitivity can be created at a specific site on the plasmid DNA containing the chicken β-globin gene when the DNA is assembled into chromatin with partially purified extracts from erythrocytes, but not from non-erythroid cells (Emerson & Felsenfeld, 1984). Using the technique of exonuclease III protection, it has been shown that the 5′- flanking regions of the *Drosophila hsp* 70 and *hsp* 83 genes make complexes with two specific proteins at their TATA boxes and the regions between -40 and -108 of the genes (Wu, 1984a,b). Moreover, a double-stranded DNA binding protein, which can recognise a sequence of dyad symmetry TGGCANNNTGCCA, has been identified. This protein was shown to bind at three sites upstream of the chicken lysozyme gene, one of which corresponds to the DNase I hypersensitive site (Borgmeyer et al., 1984). Taking these observations into

consideration together, DNase I hypersensitive sites appear to be formed around an active or potentially active gene during development by association of specific proteins at particular DNA sequences where specific structures may be formed.

As described previously (see Chapter 1.3), the *D. melanogaster Adh* gene produces two different transcripts in a stage specific manner during development. These transcripts differ only in the 5'- untranslated sequences. I have demonstrated in previous experiments (see Chapter 3.2) that DNA sequences in the *Adh* gene region possess particular conformational properties and that DNA structures plastically alter under torsional stress or microenvironmental change. It is of particular interest to know whether DNase I hypersensitive sites around the *Adh* gene region, if any, change in accordance with the transcriptional switch from the larval to the adult promoter during development. Therefore, I analysed DNase I hypersensitive sites around the *Adh* gene region in nuclei isolated at various stages, and looked for correlation between the DNase I hypersensitivity in chromatin and the plasticity in DNA structure.

3.5.2 Materials and methods

(a) Collection of embryos, larvae, pupae and adult flies.

The wild type strain (Canton-S) of *D. melanogaster* was kept in a 10-litre bucket at 25°C. A 20 cm dish containing

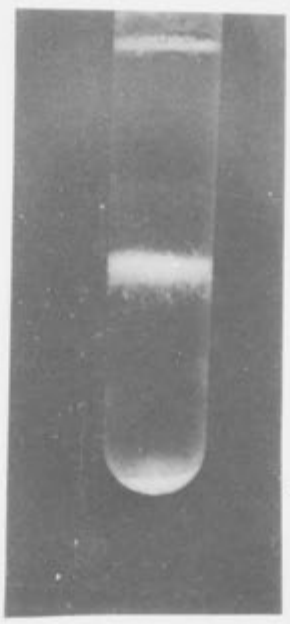
food was put in for 6 hrs to collect eggs, then the eggs laid in the dish were incubated at 25°C until they grew to appropriate stages, i.e. embryos for 1 hr, first instar larvae for 36 hrs, second instar larvae for 60 hrs, third instar larvae for 84 hrs. Larvae which were walking on the wall of the dish were collected and referred to as wandering larvae. Pupae were collected one day after pupation. A large number of larvae at each stage were collected from the dishes as follows; 20% sucrose solution was poured on the food and left for several minutes to float larvae. The floated larvae were transferred into a test tube, then an equal volume of 0.7% saline solution was layered on it. The tube was centrifuged at ~~15,000~~ 1,000 rpm for a second. Only larvae were obtained at the interface of the solution as a sharp band, while contaminating food and rubbish were precipitated and most empty egg shells floated at the top of the solution (Fig. 3.5.1). The larvae at the band were collected and washed with 0.7% saline several times, followed by freezing in liquid nitrogen. They were stored at -80°C until use.

(b) Isolation of nuclei and digestion by DNase I

The material from each stage was homogenized in approximately 6 volumes of Buffer A [0.25 M sucrose, 15 mM Tris-Cl (pH 7.4), 60 mM KCl, 15 mM NaCl, 1 mM EDTA, 0.1 mM EGTA, 0.5 mM spermidine, 0.5 mM dithiothreitol and 0.1 mM phenylmethylsulfonyl fluoride (PMSF)] in a Dounce homogenizer by 8-10 strokes. The homogenate was filtered through 75 μm pore size nylon screen and the filtrate was

Fig. 3.5.1 Isolation of large amount of larvae on sucrose cushion by centrifugation.

The floating larvae in the 20% sucrose on the food dish were transferred into a 12-ml tube. An equal volume of 0.7% saline was layered on the solution. After centrifugation at 1,000 rpm for one second, larvae were concentrated on the interface solutions, while the food and rubbish were precipitated on the bottom and the empty egg shells floated on the top.


empty
egg shells
larvae
food
& rubbish

centrifuged for 1 min at 2,000 rpm in a Sorvall HB 4 rotor. Nonidet-P40 was added to the supernatant at a final concentration of 0.2% (w/v) and the mixture was vortexed vigorously, followed by centrifugation for 10 min at 6,000 rpm in the same rotor. The crude nuclear pellet was re-suspended in Buffer A and layered on to a step gradient containing 0.85 M sucrose in Buffer B and 1.7 M sucrose in Buffer B (Buffer B is the same as Buffer A except that EDTA and EGTA are absent). The gradient was centrifuged for 20 min at 10,500 rpm in the same rotor. The nuclear pellet was re-suspended in nuclear buffer [15 mM Tris-Cl (pH 7.4), 60 mM KCl, 15 mM NaCl, 0.05 mM $CaCl_2$ 0.1 mM EDTA, 0.5 mM dithiothreitol and 0.25 M sucrose]

The purified nuclei were digested with an appropriate amount of DNase I. DNase I was diluted in a solution of 60 mM KCl , 15 mM NaCl, 100 mM $MgCl_2$, 0.1 M $CaCl_2$, 15 mM Tris-Cl (pH 7.4) and 0.5 mM dithiothreitol. Digestion was initiated by adding 10 µl of the appropriate DNase I solution to 190 µl of the nuclear suspension. The final concentrations of DNase I ranged from 5 to 40 units/ml. After incubation for 3 min at 25°C the digestion was terminated by adding 10 µl of 0.25 M EDTA. The digested nuclei were mixed with 200 µl of 2x proteinase K buffer [20 mM Tris-Cl (pH 7.8), 1.0 mM EDTA and 1% (w/v) sarkosyl] containing 200 µg/ml proteinase K and incubated at 37°C for at least 8 hrs. DNA was extracted with PCIA once and with chloroform once, followed by precipitation by ethanol in the presence of Na-acetate. The DNA pellet was re-suspended in TE buffer and treated with pancreatic

ribonuclease A at 10 µg/µl. The DNA was restricted with an appropriate enzyme as mentioned in each experiment.

3.5.3 DNase I hypersensitive sites around the *Adh* gene is defined during development

I attempted to define the pattern of DNase I hypersensitive sites on the chromatin around the *Adh* gene. First, nuclei were isolated from adult flies and digested with increasing amounts of DNase I as described in Materials and Methods. The DNA was purified, digested with Eco RI and separated by agarose gel electrophoresis. The DNA was blotted onto nitrocellulose filters and hybridized with a Pvu I - Eco RI fragment probe (see Chapter 2). This experiment enables to search an entire region from 1.2 kb upstream to 1.7 kb downstream of the *Adh* gene present in 4.7 kb Eco RI - Eco RI fragment. The results are shown in Fig. 3.5.2.

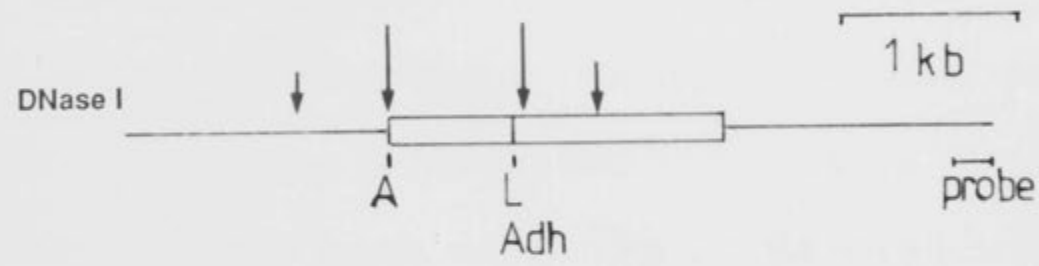
A strong 4.7 kb band corresponds to the intact Eco RI - Eco RI fragment. At higher concentrations of DNase I, the intensity of the 4.7 kb band decreased and several shorter bands, e.g. 3.5 kb and 2.7 kb, became detectable. Based on the size of detected bands, the DNase I hypersensitive sites corresponding to 3.5 kb and 2.7 kb bands were mapped at the adult capping site and at 100 bp downstream from the larval capping sites respectively. Faint 4.0 kb and 2.3 kb bands were also detectable, their corresponding hypersensitive sites were mapped about 500 bp

Fig. 3.5.2 DNase I hypersensitive sites of chromatin at the *Adh* locus in adult flies.

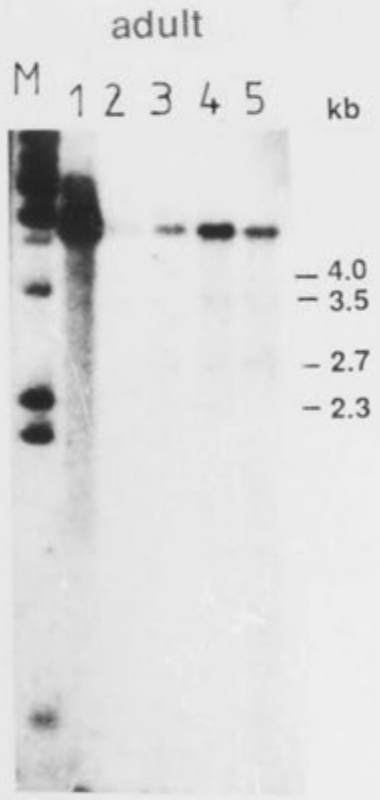
Nuclei isolated from adult flies were digested at 20°C for 3 min with DNase I at various concentrations of 0 (lane 1), 5 (lane 2), 10 (lane 3), 20 (lane 4) and 40 units/ml (lane 5), and the isolated DNA was cleaved with EcoRI, electrophoresed on 1% agarose gel followed by indirect end-labeling methods. The PvuI-EcoRI fragment was used as a hybridization probe. Size marker was a mixture of ^{32}P-end labeled HindIII-λ DNA and EcoRI-λ DNA. The detected DNase I hypersensitive sites are indicated on the schematic map of the *Adh* locus by arrows in (A).

exon.

A.



B.



upstream from the adult capping site and within the second exon.

Since the nuclei were isolated from whole fly bodies, most of the cells presumably produced no transcripts of the *Adh* gene. It appears likely that the faint bands I observed here resulted from high populations of *Adh* non-producing cells in the sample.

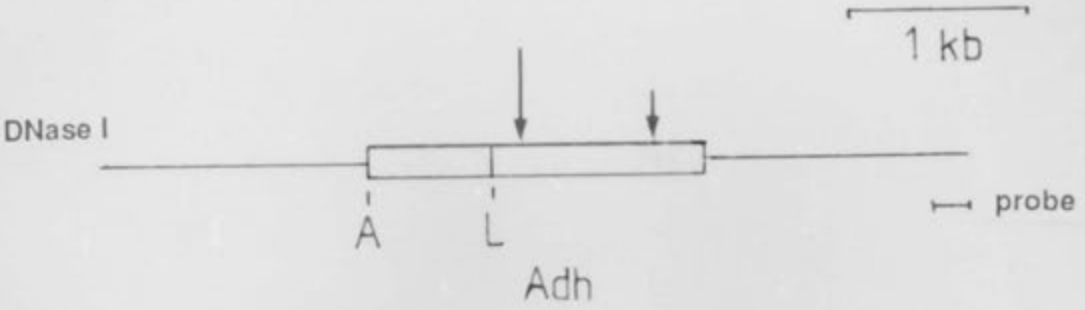
DNase I hypersensitive sites in the 4.7 kb Eco RI - Eco RI region were also analysed using nuclei isolated from the whole organism at various stages of development. A 2.65 kb band was predominantly detectable when nuclei from first instar larvae were digested with DNase I (Fig. 3.5.3). The 2.65 kb band appeared to be distinct from the 2.7 kb band in the adult fly. A 1.8 kb band was detectable but very weak. DNase I hypersensitive sites corresponding to the 2.65 kb and 1.8 kb bands were located 150 bp downstream from the larval capping site and at the 3'-region of the third exon, respectively. These two hypersensitive sites were also detectable on the digestion of nuclei from second instar larvae, and the pattern of these was almost the same as that of first instar larvae (data not shown). Moreover, the 2.65 kb band was also detectable in digestion of nuclei from third instar larvae but was rather faint. On digestion of nuclei from wandering larvae and pupae, however, no distinct band was detectable (data not shown).

All these results show that the distribution patterns

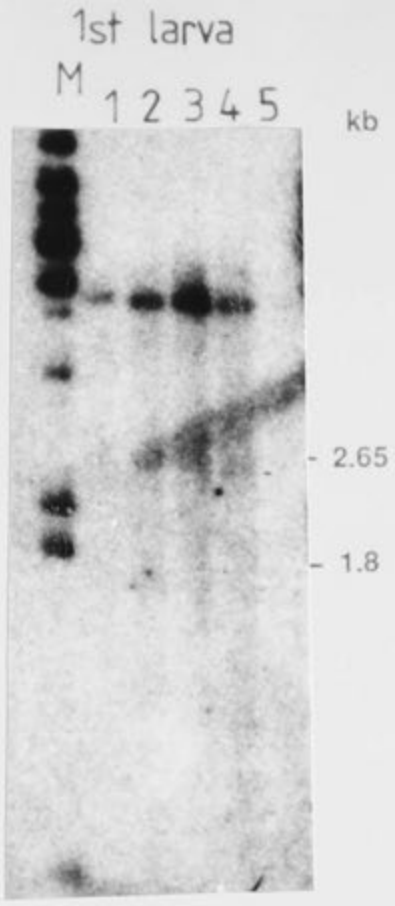
Fig. 3.5.3 DNase I hypersensitive sites of chromatin at the *Adh* locus in first instar larvae.

Nuclei isolated from first instar larvae were digested at 20°C for 3 min with DNase I at various concentrations of 0 (lane 1), 5 (lane 2), 10 (lane 3), 20 (lane 4) and 40 units/ml (lane 5), DNase I hypersensitive sites were detected as described in the legend to Fig. 3.5.2. The detected DNase I hypersensitive sites are indicated on the schematic map of the *Adh* locus by arrows in (A).

A.



B.



of DNase I hypersensitive sites in larvae are apparently different from those in adult flies. This seems to result from the differences in transcriptional activities of the *Adh* gene. In fact, the *Adh* transcripts are produced from a proximal promoter in larvae and from a distal promoter in adult flies (Benyajati et al., 1983; Savakis et al., 1986). Furthermore, no DNase I hypersensitive site in the *Adh* region was seen in pupae, implying the temporary inactivation of the *Adh* gene.

3.5.4 DNase I hypersensitive sites far away from the *Adh* gene

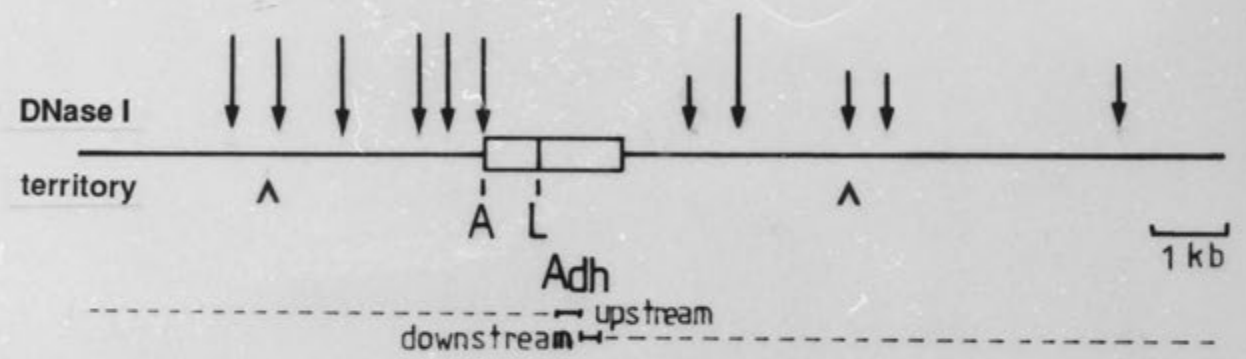
In order to analyse DNase I hypersensitive sites in the regions far upstream and far downstream from the *Adh* gene, genomic DNA isolated from DNAse I-treated nuclei were digested with Bam HI. The resulting fragments were detected by indirect end labeling methods. Bam HI upstream probe (shown in Fig. 3.5.4) was used first for detection in the far upstream region.

As shown in Fig. 3.5.4 B, several distinct bands were detected in the nuclei from adult flies. The DNase I hypersensitive sites corresponding to 1.75 and 1.25 kb bands were located about 500 bp upstream from the adult capping site and near the adult capping site, respectively. This observation confirms the result obtained in the previous experiment (see Fig. 3.5.2). Further, four bands, i.e. 4.65, 4.0, 3.15, 2.15 kb long, were also detected, the

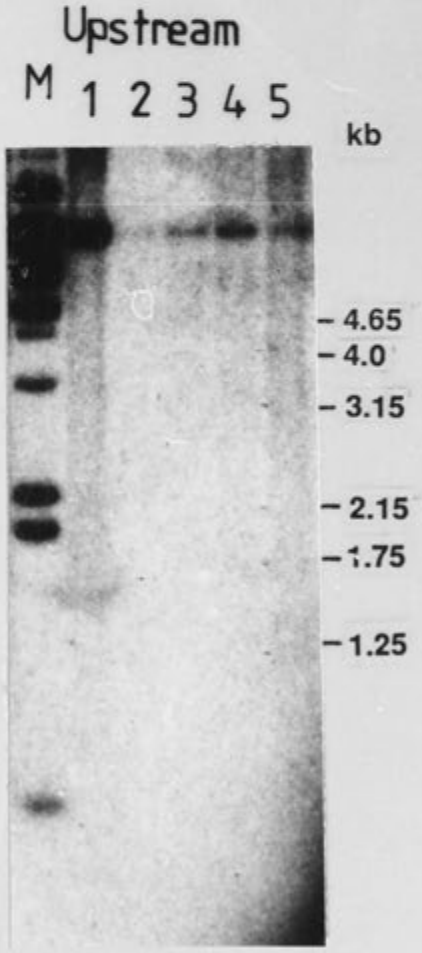
Fig. 3.5.4 DNase I hypersensitive sites of chromatin far upstream and downstream of the *Adh* gene.

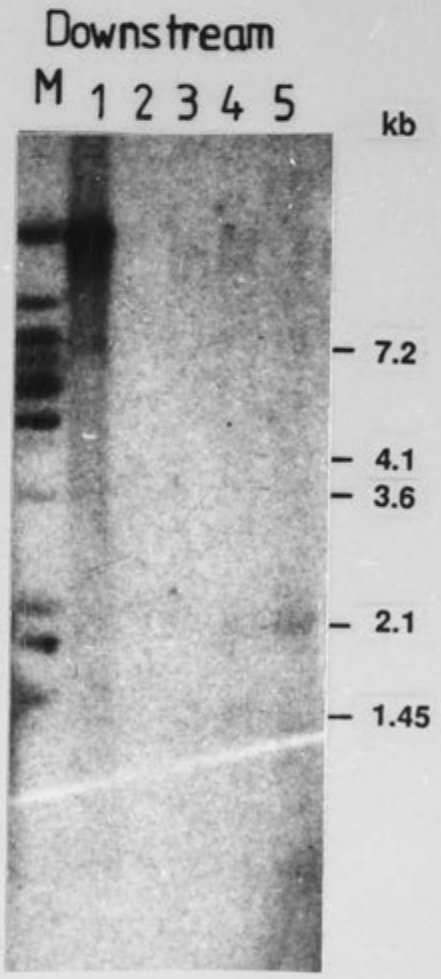
Nuclei isolated from adult flies were digested at 20°C for 3 min with DNase I at various concentrations of 0 (lane 1), 5 (lane 2), 10 (lane 3), 20 (lane 4), and 40 units/ml (lane 5). The DNA was purified, cleaved by BamHI, electrophoresed on 1% agarose, followed by indirect end-labeling methods. The BamHI-upstream (B) and BamHI-downstream (C) probes were used as hybridization probes. The detected DNase I hypersensitive sites are indicated on the schematic map of the *Adh* locus by arrows in (A).

A.


DNase I
territory
A
L
Adh
1 kb
upstream
downstream

B.


Upstream
M 1 2 3 4 5
kb
4.65
4.0
3.15
2.15
1.75
1.25


Downstream
M 1 2 3 4 5
kb
7.2
4.1
3.6
2.1
1.45

DNase I hypersensitive sites for these bands were located 3.4, 2.75, 1.9 and 0.9 kb upstream from the adult capping site (Fig. 3.5.3).

For detection in the far downstream region, Bam HI downstream probe was used. As shown in Fig. 3.5.4 C, a major band of 2.1 kb and minor bands, 7.2, 4.1, 3.6 and 1.45 kb bands, were detectable. The DNase I hypersensitive sites for these bands were located about 6.65, 3.55, 3.05, 1.55 and 0.9 kb downstream from the 3'- end of the *Adh* gene.

3.5.5 Discussion

I have examined DNase I hypersensitive sites on the chromatin (equivalent to more than 20 kb of DNA) encompassing the *D. melanogaster Adh* gene. The results showed that the DNase I hypersensitive sites were induced in a stage-specific manner during development. In cell nuclei from adult flies, several DNase I hypersensitive sites were detected. The positions of these sites are almost 3.4, 2.75, 1.9, 0.9 and 0.5 kb upstream the adult capping site, near the adult capping site, about 100 bp downstream from the larval capping site, and 0.9, 1.55, 3.05, 3.55 and 6.65 kb downstream from the 3'-end of the *Adh* gene. In contrast, on the larval chromatin, DNase I hypersensitive sites 0.5 kb upstream from the adult capping site, near the adult capping site and about 100 bp downstream from the larval capping site were lost, and two new sites were detected; 150 bp downstream from the larval

capping site and at the 3'- region of the third exon (Fig. 3.5.5). However, none of the distinct DNase I hypersensitive sites listed above were detected on the chromatin of wandering larvae and pupae.

It has been demonstrated that an actively transcribed gene is organised in an active chromatin which is predominantly susceptible to DNase I digestion at the 5'-region of the gene (Eissenberg et al., 1985). The *Adh* gene is transcriptionally silent during the period from the late stage of third instar larva to eclosion. In a larva and adult fly the different types of transcripts are produced from the dual promoter of the *Adh* gene (Benyajati et al., 1983; Savakis et al., 1986). The susceptibility to DNase I of the chromatin of the *Adh* locus seems to correlate with the transcriptional activity of the *Adh* gene. In particular, the different patterns of DNase I hypersensitive sites appears to result from the usage of two promoters of the *Adh* gene. It implies that an alteration in chromatin configuration on the *Adh* locus takes place in accordance with the switch of transcription from the larval to adult transcriptional units of the *Adh* gene. This observation, made with a dual promoter gene, suggests an interesting notion; that a change in chromatin configuration is not simply the result of transcriptional processes by RNA polymerase, but is involved in regulation of gene expression. The *Adh* gene possesses a dual promoter and a large part of the region to be transcribed is common between two transcriptional units. The small region which is not common in two transcriptional units is unlikely to

be directly responsible for a change in chromatin configuration. It seems plausible that a change in the entire chromatin configuration is required to induce an altered DNA structure at the appropriate promoter site.

In previous chapters (see Chapters 3.2 and 3.3), I analysed altered DNA structures on the *Adh* gene locus, which are induced under torsional stress. It was shown that the *Adh* gene locus exhibited a particular plasticity of DNA conformation. Some DNase I hypersensitive structures were located at the sites where altered DNA structures were induced under torsional stress. These sites are 2.75, 0.9 and 0.5 kb upstream from the adult capping site, near the adult capping site and 3.05 kb downstream from the 3'- end of the *Adh* gene. The 0.9 kb upstream DNase I hypersensitive site lies in the region of -800 to -1020, where clustered altered DNA structures are highly inducible under torsional stress, and readily recognised by a single strand binding protein (see Chapter 3.3). The 0.5 kb upstream DNase I hypersensitive site also lies in a specific region, where an altered DNA structure is induced under slightly lower torsional stress or in the presence of Mg(II) ion. This structure was also readily recognised by single strand binding protein.

Thus, it seems apparent that particular properties of DNA structure at specific sites contribute to the formation of specific chromatin structures susceptible to DNase I digestion. Since such an altered DNA structure is formed only when the DNA molecule is under torsional stress, it is

suggested that the specific chromatin structure is formed under torsional stress. This is supported by a recent experiment; the topoisomerase II inhibitor, novobiocin, removes the DNase I hypersensitivity at the site of active genes *in vivo* (Villeponteau et al., 1984). Moreover, some regulatory proteins may bind the DNA sequence at the specific site in a way similar to single strand binding protein. Other DNase I hypersensitive sites, which correspond to the locations of altered DNA structures, may also exhibit a binding preference for regulatory protein.

It should be noted that both the 2.75 kb upstream and 3.05 downstream DNase I hypersensitive sites are located at or near the borders of both 5′- and 3′- flanking territories of the adult transcriptional unit (see Chapter 3.1). This indicates that the intrinsic territories of a gene appear to be demarcated not only by altered DNA structures, but also by specific chromatin structures.

Taking all these observations into consideration, it is suggested that a dual promoter gene is organised in a characteristic domain of higher order structure of chromatin based on the plasticity of its DNA molecule. We have as yet little knowledge of mechanisms by which a DNase I hypersensitive structure is formed. However, a possible model will be discussed in Chapter 4.3.

3.6 Further implications of the territorial effects of genes: chromosomal translocation and oncogene activation

Recent observations with human Burkitt lymphomas and murine plasmacytomas have suggested that specific chromosomal translocations are linked to activation of cellular oncogenes (*c-oncs*) and these probably play a key role in oncogenic transformation (Leder et al., 1983; Yunis, 1983; Klein, 1983; Perry, 1983). For example, the human cellular *myc* oncogene (*c-myc*, the cellular oncogene homologous to the transforming gene of myelocytomatosis virus-29 (MC29)) is normally located on chromosome 8, but in some human Burkitt lymphomas the *c-myc* oncogene translocates to the immunogloblin heavy chain locus on chromosome 14, and so becomes activated (Leder et al., 1938; Yunis, 1983; Klein, 1983; Perry, 1983).

In some cases where the sequences are known, this translocation of *c-myc* is associated with primary structure alterations which may be in the untranslated first exon or in the coding exons (Rabbitts et al., 1983; Rabbitts et al., 1984; Taub et al., 1984). However, changes in coding sequences at least are not mandatory for oncogenic transformation since the only sequence changes observed in the translocated *c-myc* oncogene in the Burkitt lymphoma cell line BL22 are in the 5'- untranslated region (Leder et al.,1983).

Many models have been proposed in an attempt to

explain individual cases of activation of the translocated *c-myc* oncogene but none of them has been successful in accounting for all known cases (Robertson, 1983). Indeed Perry (1983) suggests that it may be unrealistic at this stage to attempt a unified explanation.

There has recently been proposed a hypothesis which accounts for maintenance of low levels of *c-onc* gene expression in normal cells and for enhanced expression in various forms of transformed cells (Naora et al., 1983). This hypothesis was proposed before the finding of chromosomal translocations of the *c-myc* oncogene in human Burkitt lymphomas and murine plasmacytomas (Marcu et al., 1983; Adams et al., 1983). All oncogenes may not be regulated by the same mechanism, since they are not closely related in a functional sense. However, it has been found that our hypothesis can account for most of the observations made so far on *c-onc* gene activation in human Burkitt lymphomas and murine plasmacytomas, with fewer assumptions than other models. This model is discussed here in relation to recent experiments which have linked alteration in *c-onc* gene expression to gene translocations.

The following hypothesis was based on territorial effects in a gene cluster (Naora & Deacon, 1982a & b). It had previously been noticed that when two genes (i.e. transcriptional units) lie on the same DNA strand and have an intergenic distance shorter than a defined length, the transcriptional activities of one or both genes are, in most cases, inactivated or reduced - called "territorial

effects" (Naora & Deacon, 1982a & b). It is not known at present what the *bona fide* mechanisms underlying territorial confrontation are. However, this does not at all jeopardize the observations on territorial confrontation. As previously suggested, extragenic territorial sequences may contribute in part to the formation of a higher-order chromatin structure (including a "Z" -DNA structure) which may be necessary for efficient transcription (Naora & Deacon, 1982a & b). A proposal was as follows: The transcriptional activity of the *c-onc* gene is strongly suppressed by the territorial effects of neighbouring, as yet unidentified, genes that are present within the 3′- and also probably the 5′-territories of the *c-onc* gene (Naora, Deacon & Braithwaite, 1983). In fact, recent observations show that some oncogenes or "oncogene-like" genes are located very close to transcriptionally active gene(s) on chromosomes. For example, a preliminary experiment has suggested the existence of a 5′-neighbouring gene, which is highly expressed in normal rat liver and testis cells, adjacent to the rat *c-mos* oncogene which is virtually silent in these same normal tissues (Koishi, Driver & Naora, unpublished observations). Also, in yeast DNA there is an "oncogene-like" gene, called the Yeast Protein α gene, which shows striking homology to the human *c-has/bas* oncogenes and the Harvey and Kirsten murine sarcoma viral oncogenes. This gene lies very close to the highly expressed actin gene on the same DNA strand (Gallwitz et al., 1983). In the case of human and murine *c-myc* (around 5 kbp in length (Battey et al., 1983)), these neighbouring genes are presumed to lie in a region shorter

than 13.5 kb adjacent to the oncogene (Naora et al., 1983). Any mechanisms that reinforce *c-onc* gene activity and/or reduce or eliminate the territorial effects of these putative neighbouring genes would result in enhancement of the potential activity of the *c-onc* gene (Naora et al., 1983). For example, elimination or reduction of the territorial effects of the neighbouring genes may occur as a result of deletion or damage sustained by the neighbouring genes, by enlargement of the intergenic distance between the *c-onc* gene and a neighbouring gene, or by modification of the gene's higher order conformation.

In many human Burkitt lymphomas (e.g. Joy and LSBL29) and murine plasmacytomas (e.g. M315 and S194) one *c-myc* allele has been truncated somewhere within exon 1 or intron 1 and translocated onto other chromosomes (Cory et al., 1983, Bernard et al., 1983). Of particular interest was that in these tumour cells transcripts of the untranslocated *c-myc* oncogene are formed, whereas the untranslocated *c-myc* oncogene present on normal chromosomes is virtually silent (ar-Rushdi et al., 1983; Nishikura et al., 1983; Bernard et al., 1983; Taub et al., 1984). In some cases, such as BL22 and T1033, recombination has occurred around the 5'- flanking region (within a few kb upstream from the 5'- exon) of the *c-myc* oncogene, without any break in its exons and introns (Leder et al., 1983; Corey et al., 1983). Since a complete or 5'- exon-decapitated *c-myc* oncogene is translocated within the immunoglobulin gene locus in the opposite transcriptional orientation (Cory et al., 1983; Bernard et al., 1983; Leder

et al., 1983), there should be no territorial confrontation between the translocated *c-myc* oncogene and immunoglobulin gene. Therefore, all of the above-mentioned cases can be interpreted as a simple deletion of the putative 5′-neighbouring gene, resulting in the elimination of the territorial effects of that gene. This interpretation is also applicable to cases where a variant translocation has occurred at the 3′-flanking region. For example, in the cell line IARC-BL37, which carries an 8:22 translocation, a recombination occurs between a position 400 bp 3′- to the poly(A) addition site of the *c-myc* oncogene and a position 5 kb 5′- of the J region within the λ light-chain locus (Hollis et al., 1984). This is a simple deletion of the putative 3′- neighbouring gene which was under territorial confrontation with the *c-myc* oncogene. Needless to say, an enhancer element may not be essential for activation of the newly translocated *c-myc* oncogene, but if present, as in the case of the Manca cell line (Hayday et al., 1984), it may presumably play a supplementary role in activation.

If the breakpoint is longer than 15 kb, either upstream from the 5′- exon or downstream from the 3′- exon of the *c-myc* oncogene, the translocated oncogene should still be associated with its putative neighbouring genes and consequently remain inactive as a result of the territorial effects exerted by these intact neighbouring genes. This implies that a chromosomal translocation of the *c-myc* oncogene together with intact neighbouring genes should not result in oncogenic transformation. In fact, Klein (ref. from Robertson, 1983) has mentioned that many

people have lymphocytes with typical Burkitt-type translocations and yet never develop tumours. However, it has also been noticed that the region around or within the *c-myc* oncogene is predisposed to the deletion of some sequences during or after translocation, as seen in Raji (Bernard et al., 1983; Rabbitts et al., 1983; Taub et al., 1984) and J558 (Cory et al., 1983). It is possible that a deletion of this nature may damage the function of one or both of the putative neighbouring genes, thereby resulting in the activation of the *c-myc* oncogene. Presumably this is the case with Daudi-type Burkitt lymphoma, where the breakpoint is considered to be longer than 17 kb upstream from the 5'- exon of the *c-myc* oncogene (Bernard et al., 1983).

Elevated levels of mRNA have been shown for several human Burkitt lymphomas and murine plasmacytomas which contain translocated *c-myc* oncogenes (Adams et al., 1983; Bernard et al., 1983; Marcu et al., 1983). This is, however, not always a feature of cells having a translocated *c-myc* oncogene as is seen in some cell lines, e.g. Seraphina, where no significant elevation of *c-myc* mRNA was detected (Taub et al., 1984). There may be other kinds of subtle regulatory mechanisms which still regulate *c-onc* gene expression to some extent in transformed cells. Therefore, a *c-onc* gene which has been activated by elimination or reduction of the territorial effects of neighbouring genes may be fully expressed, or may show more subtle alterations in expression, depending upon the influences of these other regulatory mechanisms. For

example, some *c-onc* genes in normal cells show cell cycle specific regulation (Kelly et al., 1983; Campisi et al., 1984), the expression of which becomes constitutive after transformation (Campisi et al., 1984; Braithwaite, Fry & LeJeune, unpublished observations). Whatever the case, however, it seems certain that *c-onc* genes that are free from territorial effects are no longer equipped with a secure mechanism to ensure against excessive gene activity. Therefore, it appears likely that these oncogenes would be more sensitive to subtle alteration in their expression.

Interestingly, this hypothesis can also account for other related observations. Firstly, recent studies of both human and murine *c-myc* oncogene sequences have shown that there are two initiation sites, located about 150-160 bp from one another, in the normal *c-myc* oncogene (Bernard et al., 1983; Battey et al., 1983 Stanton et al., 1984; Taub et al., 1984). The normal (untranslocated) gene in lymphoblastoid cells (IARC100) tends to use the second promoter, producing short transcripts (Battey et al., 1983; Leder et al.,1983; Hollis et al., 1984; Taub et al., 1984). However, in Burkitt lymphoma cells, e.g., BL22, in which the *c-myc* oncogene is translocated and has lost the putative 5'-neighbouring gene but still retains both promoters unchanged, transcription initiates from both the first and second sites (Battey et al.,1983; Leder et al., 1983; Hollis et al., 1983; Taub et al., 1984). I suggest that in normal cells initiation preferentially occurs from a site as far away as possible from the putative 5'-neighbouring gene in order to reduce the territorial

effects, whereas in BL22 both initiation sites are equally available for transcription because of the absence of any territorial confrontation at the 5′- flanking region of the translocated *c-myc* oncogene in these tumour cells.

Secondly, it has been observed with the human Burkitt lymphoma ST486 cell line that the 5′-exon of the *c-myc* oncogene remaining on human chromosome 8 after translocation becomes transcriptionally active and that abnormal transcripts are produced (ar-Rushdi et al., 1983). Since such a truncated 5′-exon of the *c-myc* oncogene is assumed to be no longer under territorial confrontation with the putative 3′- neighbouring gene, it would be actively transcribed. Once again, this would be due to simple elimination of territorial effects.

Finally, I would like to mention specific activation of the *c-mos* oncogene (cellular sequences homologous to the transforming sequences of Moloney murine sarcoma virus) in a mouse plasmacytoma (Canaani et al., 1983) and a few other *c-onc* genes in various types of tumours. In normal mouse cells the *c-mos*, approximately 1.2 kb in length, (Oskarsson et al., 1980), is virtually inactive (Muller et al., 1982)., whereas the *c-mos* oncogene in mouse plasmacytoma XRPC24, containing an extra 4.7 kb DNA element within its coding region in a 5′-5′ configuration, is transcriptionally active (Canaani et al., 1983). As the result of an insertion, the *c-mos* oncogene is activated and transcription presumably initiates at some sequences of the LTR adjacent to the main portion of the oncogene or at a

in a manner similar to those observed on the *c-myc* genes in some plasmacytoma and Burkitt lymphoma cells.

I suggest that the proviral insertion may provoke the *c-mos* activation by the elimination of territorial effects.

cryptic site within the coding region. An interpretation is that the short intergenic region between the putative 5′- neighbouring gene and the *c-mos* oncogene present on the normal chromosome is enlarged by the insertion of the 4.7 kb DNA sequence in the 5′-5′ configuration. In fact, the 4.7 kb extra DNA sequence is long enough to eliminate the territorial effect of the putative neighbouring gene from the *c-mos* oncogene. Therefore, the rearranged *c-mos* oncogene should be no longer under territorial confrontation with the putative 5′- neighbouring gene. Since the LTR sequences are inserted in the opposite orientation, LTR insertion may also influence the transcription of the *c-mos* oncogene. In relation to *c-mos* activation by LTR insertion, it should be noted that a sequence of DNA (called "UMS") present about 1 kb upstream from the mouse *c-mos* oncogene suppresses activation of the transforming region of the *c-mos* or *v-mos* oncogenes (Wood et al., 1983). These suppressive effects of the UMS are somewhat similar to the territorial effects of flanking genes. Since a transcriptional unit, highly active in normal cells, is predicted to be located in almost the same region as the mouse UMS from our hypothesis, the UMS could contain an active transcriptional unit. If so, the effects of the UMS should be territorial suppression, and thus these observations provide strong support for our hypothesis. More experiments are needed to clarify this point.

A similar interpretation may be possible in the case of *c-myc* activation in a T-cell lymphoma (Adams, 1984) and

the *int-1* oncogene in some of the mouse mammary carcinomas (Nusse et al., 1984). In these cases, proviral DNA sequences are inserted in a 5′-5′ configuration around the proximal region of the *c-onc* gene, thereby presumably enlarging the intergenic distance on the same DNA strand. Once again, this implies elimination of the territorial effects of the neighbouring gene.

Thus, most of the observations made with human Burkitt lymphomas and murine plasmacytomas can be explained with the gene territoriality hypothesis, provided that there are transcriptionally active genes located near the *c-myc* oncogene. The *c-mos* oncogene is one example of where a neighbouring transcriptional unit is present, and when active, suppresses *c-onc* gene transcription. To date, however, there is no evidence that the *c-myc* oncogene lies near to other active genes in its normal configuration. I predict, then, the existence of *transcriptionally active genes in a normal cell* in close proximity (within 13.5 kb) to the *c-myc* oncogene.

3.7 A neighbouring gene actually exists near the *D. melanogaster Adh* gene as predicted by the gene territoriality hypothesis

3.7.1 Introduction

The gene territoriality hypothesis has been already described in detail in Chapters 1.1 and 3.1. Probably, extragenic territorial DNA sequences are required not only for clustered genes but also for non-clustered genes. It seems that this hypothesis comprehensively accounts for various observations about gene expression in normal and cancer cells. My proposal is that gene territories have an important role in the regulation of gene expression in eukaryotes during development.

The *D. melanogaster Adh* gene is transcribed from different capping sites at different specific stages of development as already mentioned (see Chapter 1.3). Recently, however, it has been reported that *D. mulleri* (a distant relative of *D. melanogaster*) produces two different ADH proteins, i.e. ADH-1 (larval) and ADH-2 (adult), in a temporal pattern similar to the expression of the *D. melanogaster Adh* larval and adult transcripts (Batterham et al., 1983). Consequently, it has been found that the two proteins of *D. mulleri* are encoded by two distinct genes separated from each other by 2 kb of DNA, *Adh*-1 and *Adh*-2 (Fisher & Maniatis, 1985). As suggested by P-element transformation of *D. melanogaster* with the *D. mulleri Adh* genes, some separate regulatory elements for transcription

may mediate the specificity of *Adh* expression in different specific tissues and stages (Fisher & Maniatis, 1986). In the transformants, however, the expression patterns of the introduced *Adh* genes were not identical to those of the *D. melanogaster Adh* gene. Further, the 5′- flanking regions of the *Adh*-1 and *Adh*-2 genes show little sequence homology with the DNA nucleotide sequences present near the larval and adult promoter of the *D. melanogaster Adh* gene (Fisher & Maniatis, 1985). Thus, it seems that regulatory elements alone do not sufficiently explain the differential expression of the larval and adult *Adh* genes during development.

I predict that a putative neighbouring gene exists at the 5′- or 3′- flanking region of the *D. melanogaster Adh* gene. The putative gene should be close enough to the *Adh* gene to bring about territorial effects on the distal promoter, but not on the proximal promoter, for the longer transcriptional unit from the distal promoter requires the larger territory. According to the gene territoriality hypothesis, moreover, the putative gene, if present, should be actively transcribed during larval development when the distal promoter is inactive. Conversely, it should become inactivated in adult flies where the distal promoter becomes active.

In this experiment, I have examined the possible regulation of the differential expression of the *Adh* gene through gene territoriality. I have attempted to answer whether the putative gene actually exists in the 5′- or 3′-

flanking region of the *Adh* gene.

3.7.2. Materials and methods

(a) RNA preparation

Total RNAs were prepared from embryos, larvae, pupae and adult flies by the guanidine thiocyanate-CsCl method (Kaplan et al., 1979). Frozen material was thawed and homogenized with RNA extraction buffer [5 M guanidine thiocyanate, 10 mM EDTA, 50 mM Tris-Cl (pH 7.6), 5% (v/v) 2-mercaptoethanol] in a Dounce homogenizer with about 10 strokes, then N-lauryl sarcosine was added to the homogenate at a final concentration of 4% (w/v) and mixed well. CsCl was added to the mixture at 0.5g per millilitre, and dissolved well. The mixture was layered on a cushion of 5.7 M CsCl and centrifuged for 20-24 hr at 20°C in an SW50.1 rotor at 35,000 rpm. The RNA pellet was washed with 70% (v/v) ethanol and dissolved with 0.25% N-lauryl sarcosine. The RNA solution was treated with phenol twice, precipitated with sodium acetate and ethanol. The RNA pellet was dissolved in 0.25% N-lauroylsarcosine and stored at 20°C.

Poly(A) RNA was purified from second instar larval RNA using mAP (Amersham) according to the published procedure.

(b) Southern hybridization with c-DNA probe

To label a probe for hybridization, second instar

larval poly(A) RNA was transcribed by AMV reverse transcriptase in 50 µl in RT buffer [50 mM Tris-Cl (pH 8.5), 40 mM KCl, 10 mM $MgCl_2$, 0.4 mM dithiothreitol, 800 µM each of dCTP, dGTP and TTP, 40 µM dATP, 20 µCi [α-^{32}P]-dATP, 5 µg/ml hexa-deoxynucleotide and 30 units of AMV reverse transcriptase] for 60 min at 42°C. After RNA templates were removed by incubation in 0.3 M NaOH for 12 hr at 37°C, the labelled DNA was precipitated with NH_4-acetate and ethanol.

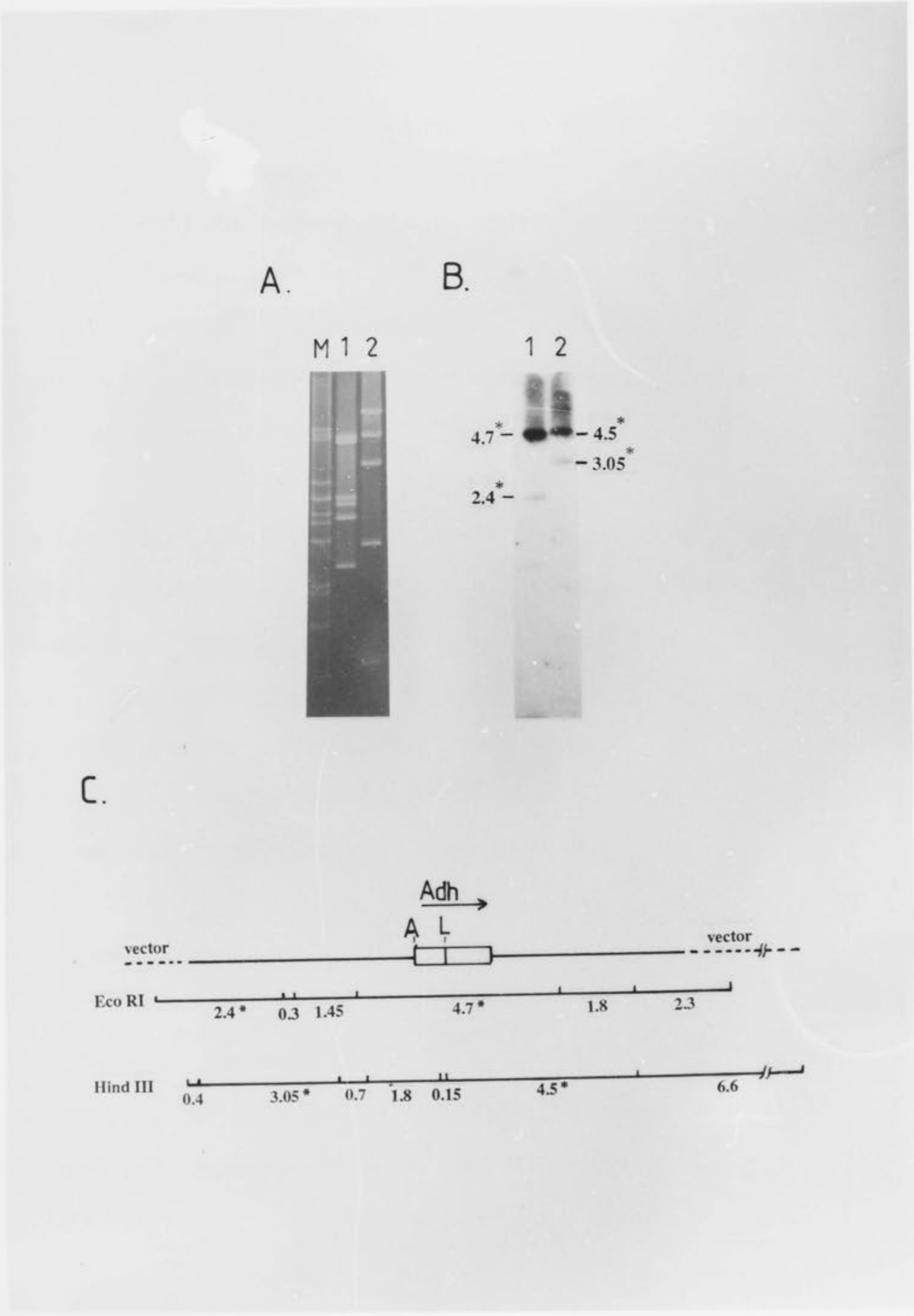
Plasmid DNA, psAF2 was digested with Hind III and Eco RI respectively, and run on a 1% (w/v)agarose gel in TAE buffer followed by Southern hybridization with ^{32}P-cDNA probe from second instar larval poly(A) RNA.

3.7.3 A putative neighbouring gene close to the *D. melanogaster Adh* gene

On the basis of the gene territoriality hypothesis, I predicted that a neighbouring gene exists close to the *D. melanogaster Adh* gene. This prediction was examined here using a recombinant DNA, psAF2, which contains an 11.8 kb Sac I fragment including the *Adh* gene and both its flanking regions (Goldberg, 1980; Fig. 3.7.1 A). Eco RI and Hind III restriction fragments of psAF2 DNA were Southern-transferred onto a nitrocellular filter, and the filter was hybridized with ^{32}P-labelled cDNA transcribed by reverse transcriptase from poly(A) RNA, which was isolated from larvae of *D. melanogaster*. Transcripts of the putative

Fig. 3.7.1 Detection of a neighbouring gene of the *Adh* gene by hybridization with cDNA.

The EcoRI- (lane 1) and HindIII- (lane 2) digests of psAF2 DNA were electrophoresed on 1% agarose and transferred to nitrocellulose filter. The pattern of the ethidium bromide-stained gel is shown in (A). The size marker is a PstI-λ DNA. The blotted filter was then hybridized with ^{32}P-labeled cDNA prepared from poly(A)-RNA purified from larvae (B). The restriction map of the *Adh* locus in psAF2 plasmid DNA is shown in (C). The fragments with asterisks (*) represent the fragments detected on cDNA hybridization.


A.
B.
M 1 2
1 2
4.7*
4.5*
3.05*
2.4*
C.
Adh
A L
vector
vector
Eco RI
2.4 *
0.3
1.45
4.7 *
1.8
2.3
Hind III
0.4
3.05 *
0.7
1.8
0.15
4.5 *
6.6

neighbouring gene were predicted to be produced during larval development when the adult transcriptional unit was under territorial confrontation with the putative gene. The result of the hybridization is shown in Fig. 3.7.1. The 4.7 kb Eco RI and 4.5 kb Hind III fragments were strongly detected and the 1.8 kb Hind III fragment was faintly detectable. This observation indicates that the proximal transcripts of the *Adh* gene are actually produced as a poly(A) RNA in the larvae, but the distal transcripts are not. This is in accordance with other observations by Northern-blotting as before and by other authors (Benyajati et al., 1983; Savakis et al., 1986). Thus the hybridization experiment of restriction fragments with ^{32}P-labelled cDNA probe seems to be a useful method to detect which region of a closed DNA is transcribed.

In the hybridization, furthermore, 3.0 kb Hind III and 2.4 kb Eco RI fragments were also detectable though they do not encode any part of the *Adh* gene (Fig. 3.7.1). These findings indicate that DNA sequences at least in the Hind III - Eco RI fragment about 3 kb upstream the *Adh* gene are transcribed as poly(A) RNA in the larvae as actively as the larval transcriptional unit of the *Adh* gene.

3.7.4 Discussion

I predicted the existence of a putative neighbouring gene in the 5′- or 3′- flanking region to the *Adh* gene on the base of the gene territoriality hypothesis.

Furthermore, the putative gene should be located in a limited region to have territorial effects upon the adult promoter only, not the larval promoter.

I have demonstrated that the putative gene actually exists in the 5′-flanking region to the *Adh* gene. Although I have not determined the exact location of the gene, I showed that at least a part of the transcribed region of the gene lay about 3 kb upstream from the *Adh* gene. It indicates that the intergenic distance between the putative gene, and the adult or larval transcription unit of the *Adh* gene is about 3 or 3.7 kb long. The curve of relationship between total gene sizes and intergenic distances in Fig. 3.1.1 indicates that the total gene size should be 1.8 kb to 2.7 kb. In other words, the size of the putative gene is 0.1 to 1.1 kb. I have not determined the size yet, but I predict a size of about 1.1 kb for the neighbouring gene.

The neighbouring gene appears to be actively transcribed in larvae. Therefore, it is apparent that the putative gene and the larval unit are not under territorial confrontation. Probably the adult unit is under territorial confrontation inactivated at this stage. In contrast, the adult unit is actively transcibed in an adult fly, where the putative gene is predicted to be under territorial confrontation and to be inactivated. Thus, the expression of the dual promoter gene, may be tightly related to that of the neighbouring gene.

C H A P T E R F O U R

GENERAL DISCUSSION

4.1 DNA sequences are organised in a special way for gene functions

It has been suggested that gene families evolved by duplication followed by conservation and variation of family members (e.g., Efstratiadis et al., 1980). During a long period of evolution, regulatory elements such as CAAT- and TATA-boxes, and AAATAA are well conserved and their positions are defined (Proudfoot & Brownlee, 1976; Benoist & Chambon, 1981; Breathnach & Chambon, 1981; Dierks et al., 1983). Other tissue-specific elements for transcription have been also revealed by DNA sequence analysis (Garfinkel et al., 1983; Pelham, 1982 and Fujita et al., 1985).

In the members of the *Drosophila Adh* gene family, most of the 5'- flanking regions show little significant sequence conservation except for the above mentioned regulatory elements (Fisher & Maniatis, 1985). I have compared the far upstream 5'- flanking region of the *D. melanogaster Adh* gene with those *D. mulleri* genes, and again showed little sequence homology among them. Thus, the primary structure or nucleotide sequence of extragenic regions does not seem to be conserved in the organization of the *Adh* genes.

In contrast, the secondary structure of the DNA molecule does appear to be conserved. This notion is supported by several observations.

(i) The distribution patterns of free energy of base pairing are very similar among all the *Adh* genes (Chapter 3.2). In this experiment, I have shown that the protein-coding sequences exhibit a relative low free energy or are thermodynamically stable while the extragenic sequences exhibit a high free energy. A striking observation is that the free energy is apparently high in the introns and 3'-untranslated regions (Chapter 3.3). Thus, DNA sequence seems to be organised in concert with thermodynamical stability and its organization appears to be conserved among various *Adh* genes of *Drosophila*. Such a organization of DNA sequence is specifically recognized by a single strand binding protein when the molecule is under torsional stress (Chapter 3.2). Therefore, thermodynamic stability would be readily revealed by negative torsion. All these observations suggest that precise nucleotide sequences of extragenic and intron sequences are not conserved but the overall situation of the thermodynamical stability is conserved.

Similar properties of DNA sequence organization have been revealed in the heat shock locus by micrococal nuclease digestion (Keene & Elgin, 1981; 1984). In this case, non-transcribed regions tend to have preferential cleavage sites while transcribed regions are relatively free of such prominent cleavage sites. Furthermore, as the prominent sites are frequently spaced 200 bp apart, their distribution is suggested to bear a functional relationship to the nucleosome organization.

(ii) Altered DNA structures have been detected at or near the 5′ end of several genes by S1 nuclease digestion (Larsen & Weintraub, 1982; Selleck et al., 1984; Kilpatrick et al., 1986). I have also found such altered structures at 5′-ends of both adult and larval transcription units, respectively. In addition to these structures, around 500 bp upstream from the 5′- ends of the adult and larval units, other peculiar altered structures are located which can be detected at slightly low superhelical density by S1 nuclease (chapter 3.2). It is apparent, therefore, that there is a pair of altered structures in the 5′-flanking regions of respective transcriptional units, i.e. one is formed at high torsional stress, and the other is at slightly lower torsional stress. A pair of altered structures may be required for transcriptional regulation of each unit. Once again, this observation indicates that organization of altered DNA structures is well conserved in the extragenic regions of two transcriptional units present in a dual promoter gene although no conservation of their nucleotide sequences has taken place in these regions. It is possible that the altered DNA structure present at 750 bp upstream from the adult capping site might have a common role in transcriptional regulation of both units. These assumptions may be examined by P element transformation using an *in vitro* mutated *Adh* gene.

(iii) I have shown that the intrinsic territories of a gene are demarcated by the presence of a specific altered DNA structure. In the case of *D. melanogaster Adh* genes, for example, they are 2.7 kb and 0.75 kb upstream from the

adult capping site, and 0.8 kb and 3.0 kb downstream from the poly(A) addition site (Chapter 3.1). Similarily such altered structures are observed at the putative borders of intrinsic territories in other genes. Albeit sequences diverge at these sites, locations of altered DNA structures seem to be well defined (see also Chapter 3.1).

(iv) Other types of organization of DNA sequence have been reported to be conserved. First, the segments partially denatured in 80% formamide (presumably AT-rich sequences) have been shown to punctuate the genomic DNA systematically (Moreau et al., 1981). In the chicken α- and β- globin gene families, such segments frame the gene domains (Moreau et al., 1982). It has been shown that this pattern is similar between the α- globin families of chicken and duck (Kretsovali et al., 1986). Thus, such a large scale organization of DNA sequence also seems to be conserved. Secondly, some of the nucleotides have a clear tendency to be periodically repeated at about 10.5 bases along the eukaryotic DNA sequences. As this periodicity is equal to the pitch of chromatin DNA, the above observation implies that the deformational anisotropy of the DNA molecule at the dinucleotides facilitates its smooth folding in chromatin (Trifonov & Sussman, 1980; Trifonov; 1980).

Thus all these facts suggest that the genomic DNA in eukaryotes is organized at the secondary structure or higher order structure level on a large scale in a special way. A secondary structure of DNA duplex rather than a

sequence itself is likely to play some functional roles. For example, such organization may be important not only in satisfying the basic requirements for regulatory protein recognition and binding at a specific site, but also in determining the overall conformation of higher order structures in the DNA molecule outside and within a structural gene.

4.2 DNA conformation can plastically change under torsional stress

The conformational features of DNA molecules around the *D. melanogaster Adh* gene region have been investigated by nuclease digestion of recombinant DNA containing the gene (Chapter 3.2). There are several altered DNA structures at specific sites encompassing the *Adh* gene, which can be formed under torsional stress. At least three types of altered DNA structures have been characterised; (1) The structure mapped at about 750 bp upstream from the adult capping site. It is strongly induced according to the increase of superhelical density and recognized by various types of nuclease in respective pH conditions. It is also stabilized by Mg(II) ion. (2) The structures mapped at about 500 bp upstream from the adult and larval capping sites respectively. They are not detectable simply according to the increase of superhelical density, but it is formed at a limited level of superhelical density. They are also induced and stabilized by Mg(II) ion. (3) The structure mapped at about 800 bp downstream from the 3′ end of the *Adh* gene. The formation of this structure also

proportionally depends on the superhelical density. However, it is eliminated by Mg(II) ion.

A question is raised as to why the second type structure is not detected at high superhelical density. It is possible that the formation of a certain structure or structures prevents the detection of others. This striking phenomenon could be accounted for by the two following mechanisms; (1) The formation of an altered DNA structure at one site precludes the establishment of others by competing for torsional energy. Alternatively, (2) altered DNA structures at other sites are actually formed but undetectable as the structure of the primary site is preferentially recognized, because the torsional stress will be lost and altered DNA structure will disappear immediately after a nick is introduced at the primary site. In any case, it seems certain that the second type structure has a stable conformation at a lower superhelical density than the first type structure. A number of unusual features of nucleotide sequence that can affect the local conformation of DNA under torsional stress have been detected by S1 nuclease attack.

The nucleotide sequences around the positions of altered DNA structures of all the types indicate that these nucleotide sequences responsible for altered structures are different from those reported so far (Chapter 3.3). Although their actual conformations are unknown, it is possible that a variant structure like bending is deformed

further by torsional energy. Most notably, a detailed mapping of S1 attack sites has shown that a fine structure at the -500 site alters depending on the superhelical density, but is not gradually formed according to the increase of superhelical density. The observation indicates that an altered structure and a normal B-form structure are not alternatively formed at a threshold of superhelical density. Even in a small region, torsional stress can dynamically change the conformation of a fine structure depending on the strength of torsion.

Thus the different behaviours of these structures are displayed according to the superhelicity and ionic conditions. It is unlikely that the sequences at which the altered DNA structures are induced under torsional stress would simply behave as a sink of torsional energy. Considerable torsional energy can be stored throughout an entire circular DNA. It seems plausible that the altered DNA structures detected around the *Adh* gene region will interact with one another over a long distance in the circular DNA to compete for the torsional energy. Such a long range interaction can be dynamically modulated by microenvironment, e.g. ionic condition. The interaction would not necessarily require any direct physical contacts of the structures. Two types of topoisomerase can change the superhelicity of circular DNA molecule. Thus, the interaction of the altered DNA structures could be dynamically modulated in an indirect manner if the topoisomerase binds any other sites and changes the superhelical density. Another modulation of the

interaction seems to be possible, too. For example, if any regulatory proteins recognized and bound the altered DNA structures and thus the bound structures became unavailable for mutual interaction, then dynamic interactions could only take place with other available structures in a superhelical domain.

Moreover, I have demonstrated that a single strand binding protein (SSB) can modulate such an interaction under torsional stress (Chapter 3.2). The similarity of the distribution patterns between free energy of thermodynamic stability of DNA duplex and SSB binding implies that SSB would preferentially bind a relatively unstable region of DNA duplex (Chapter 3.3). However, SSB does not seem to be able to bind directly to such an unstable region of DNA duplex. The binding of SSB requires a negative supercoiling which produces a primer site of single strand, because SSB can not bind relaxed circular DNA (Chapter 3.2). SSB can recognize a single strand in an altered DNA structure and make a denaturation bubble by binding the single strand (Kowalczykowski et al., 1981; Glikin et al., 1983). At increased concentration of SSB, however, SSB can not fix the denaturation bubble at the position where the altered DNA structures are originally formed (Chapter 3.2). The formation of the denaturation bubble reduces the torsional energy by fixation of the unwound strand. Consequently, other altered DNA structures at other sites newly appear at a reduced torsional energy. Thus, structures induced by torsional stress can mutually interact with others.

However, SSB will eventually fix denaturation bubbles in thermodynamically unstable regions of a DNA duplex at the high concentration of SSB. This is my view how the modulation of topological state and conformational interaction through torsional energy in a circular DNA molecule takes place on chromosomes.

The dynamic interaction is limited within a circular DNA molecule, and never occurs outside the circle. It implies that a circular DNA molecule will behave as a topological domain. When both ends of duplex DNA molecule are rotationally fixed, the entire molecule will become a topological domain as a "circular" form. It should be mentioned again that the torsional energy is stored throughout the entire circular DNA molecule. Thus such a DNA molecule behaves as a topological domain in which dynamic conformational interaction can occur without direct physical contact.

Finally, such a dynamic interaction reminds us of allosteric proteins. Allosteric proteins can dynamically change their whole conformation to acquire different functions by the binding of chemical agents or the contact of other proteins. It is well known that proteins are also conformationally quite flexible in solution according to ionic conditions and pH. It implies that biological macromolecules must have plasticity in nature, which allows them to play a dynamic function in complex and fine regulations.

4.3 Higher order structure and transcription

One of the interesting observations I made in my experiments is that the DNase I hypersensitive sites on chromatin around the *Adh* gene alter in concert with the transcriptional activities of the larval and adult transcriptional units during development (Chapter 3.5). This observation with the *D. melanogaster Adh* gene is the first case in a dual promoter gene. The change of chromatin structures is not the result of the transcriptional process of RNA polymerase. This suggests that a higher order structure of chromatin is deeply involved in the mechanisms of transcriptional regulation of a dual promoter gene.

Most chromosomes in eukaryotic nuclei are condensed into a packed structure. RNA polymerase and other transcriptional machineries may not be readily able to interact with DNA organised in such a manner. In fact, the cloned *D. melanogaster Adh* gene can be transcribed *in vitro* in the cell free extracts prepared from various phylogenetically diverse sources (Chapter 3.3). Moreover, the exogenous *Adh* genes in transfected cell are readily transcribed, while the endogenous genes, which are in a different chromosomal position, are not (Benyajati & Dray, 1984). Therefore, when a gene on a chromosome is transcribed, the DNA molecule of the gene should be rearranged in a specific structure, which exhibits a free accessibility to transcriptional machineries. My observations in the *in vitro* transcription experiments

further indicate that the transcriptional machineries are universal in various phylogenetically diverse animals (Chapter 3.3). Therefore, the regulation of higher order structure of chromatin must have an important role in the tissue- or stage- specificity of transcription.

As indicated in Chapter 3.5, some DNase I hypersensitive sites in chromatin at the *Adh* locus correspond to the sites of altered DNA structures induced by torsional stress. We have little knowledge about the molecular mechanism underlying the formation of DNase I hypersensitive structures. However, it is possible that a DNA secondary structure itself or its potential properties are used to create a DNase I hypersensitive structure in chromatin. For example, some of the altered DNA structures may be directly involved in physical curving or bending of chromatin or a specific plasticity of DNA may lead to local variation of chromatin structure.

Some altered DNA structures may also represent a general feature of DNA sequences recognized by a specific class of regulatory components. A protein-DNA complex may form on chromatin, which is susceptible to DNase I attack. Other DNA structures could also be involved in nucleosome phasing. Assuming that the DNA is folded around a nucleosomal core in a peculiar way, it could exert its effect over a distance of 200 bp in chromatin. In fact, it has been recently reported that a DNA structure preferentially exists in a linker region of nucleosome (Nickol & Martin, 1983).

The chromatin fibres are arranged into a series of looped structures in association with nuclear matrix (Paulson & Laemmli, 1977; Vogelstein et al., 1980; Mirkovitch et al., 1984). In Chapter 4.2, I described the possible mutual interaction of altered DNA structures in a circular DNA molecule. It is highly likely that a similar dynamic interaction in a loop domain takes place on chromatin, based on the plasticity of a naked DNA molecule. This would provide us with an interesting idea as to how the expression of genes on chromatin is regulated; (1) A modulation of gene expression can be achieved by supercoiling of different regions of the genome to different extents. (2) Several genes in a loop domain can be coordinately regulated through changes in superhelical density of its loop. (3) The length of such a loop domain can be altered by changing the anchorage sites of a chromatin fibre to nuclear matrix. This is actually observed during the period of a cell cycle and at developmental stages (McCready et al., 1980; Vogelstein et al., 1980; Buongiono-Nardelli et al., 1982).

Little is known regarding the actual molecular mechanism by which torsional stress affects the transcriptional activity of genes. If torsional stress forces certain DNA sequences free from a repressive chromatin structure, a chromatin structure within a supercoiled domain of loops will be maintained in an energetically high state. Such a chromatin structure can dynamically alter not only by torsional stress in the domain, but also by microenvironmental factors, e.g. concentrations of metal ions, salt concentration and

hydrophobicity (Saenger, 1983; McGhee & Felsenfeld, 1980). Dynamic alteration of chromatin structures in these cases seems to depend upon the plasticity of supercoiled DNA.

It should be mentioned here that the plasticity of supercoiled DNA may even be directly involved in a mechanism underlying differentiation of two daughter cells during development. For example, it is widely held that various kinds of gradients, e.g. salt or other metal ions, play an important role in developmental determination, commitment and differentiation in animal embryos and cultured cells (Davidson, 1976). I speculate, therefore, that certain DNA sequences present in a particular supercoiled domain of chromatin may display a specific chromatin structure corresponding to the particular microenvironment in a certain gradient, whereas the same DNA sequences may display another structure elsewhere in the same gradient. Therefore, when two daughter cells are exposed to different parts of the gradients, the chromatin structures of corresponding genes in the two cells would be different, thereby leading to differentiation of cell functions in different cell lineages.

I have demonstrated that the transcriptional activity is enhanced according to the increase of the superhelical density in the *Adh* gene contained in recombinant circular DNA (Chapter 3.4). It is not likely that torsional stress induced an active higher order structure of chromatin during *in vitro* incubation since the reaction time was too short to form it. RNA polymerase possesses a high affinity

to single stranded DNA. In the *in vitro* reaction mixture, therefore, RNA polymerase may recognize a single stranded region in an altered structure induced by torsional stress and then move to the initiation sites along the DNA molecule, followed by transcription of the coding region. Alternatively, regulatory proteins for transcription can recognize an altered DNA structure and form an active protein-DNA complex. In either case, however, an interesting mechanism is involved in the enhanced transcription in that the regulation of torsional stress by topoisomerase or other factors would result in a fine control of the transcriptional activity of a gene in a superhelical domain.

4.4 Gene Territoriality and coordinated regulation

I predicted the existence of a putative neighbouring gene in the 5′- and/or 3′- flanking region to the *Adh* gene on the basis of the gene territoriality hypothesis (see Chapter 3.1). As predicted, I have demonstrated here that the putative gene indeed lies about 3 kb upstream from the *Adh* gene (Chapter 3.7). This newly discovered gene appears to be actively transcribed at the larval stage. Although I have not characterised the neighbouring gene in detail, I predict from the gene territoriality hypothesis that the size of the gene is about 1.1 kb assuming that the expression of the neighbouring gene is related to the alternative use of the dual promoters of the *Adh* gene. In this situation, the longer transcript from the adult promoter should not be produced simultaneously with the

transcript from the neighbouring gene, mainly due to the the territorial confrontation between the neighbouring gene and the adult transcription unit of the *Adh* gene.

Similar explanation has been made for the transcriptional activity of another dual promoter gene, the *c-myc* oncogene (see Chapter 3.6). In most Burkitt lymphoma cells, the distal promoter of the *c-myc* gene is preferentially active, whereas the proximal promoter is predominant in normal cells (Battey et al., 1983; Hollis et al., 1984; Taub et al., 1984). The gene territoriality hypothesis accounts for the preferential usage of the distal *c-myc* promoter in cells with translocations. In a normal cell, a putative 5'-neighbouring gene, predicted from the hypothesis, is supposed to reduce the transcription from the distal *c-myc* promoter by territorial effects since the transcriptional unit from the distal promoter requires a larger territory than that of the proximal promoter. In Burkitt lymphoma cells, both promoters are free from the territorial confrontation of the 5'-flanking gene as a result of chromosomal translocation.

Thus, it seems reasonable to propose that switching of transcriptional activities of a dual promoter gene is regulated in conjunction with the expression of a neighbouring gene. Apparently, both promoters of a dual promoter gene is never utilized for their transcription at the same time. How is their expression coordinately regulated?

In several eukaryotic systems, it has been found that transcripts extending beyond the 3'-end of the mature mRNA are synthesized during transcription (Hoffer & Darnell, 1981). This observation may provide us with an explanation for the inhibition of utilization of a downstream promoter in the same orientation to a upstream promoter (Cullen et al., 1984; Proudfoot 1986). Similar inhibition has also been observed in prokaryotes as "promoter occlusion"; the upstream promoter of prophage λ inhibits the initiation of transcription at the downstream *gal* operon promoter (Adya & Gottesman, 1982). It is possible that such transcriptional interference occurs at the *Adh* locus. However, it can not fully account for the coordinated regulation of *Adh* gene expression, because the distal promoter of the *Adh* gene is used when both the upstream neighbouring gene and the proximal promoter are silent.

Moreover, in the *E. coli* lactose operon, an upstream promoter is located at position -22 to the principal promoter. It has been observed in the operon that the addition of cAMP receptor protein and cAMP results in the repression of the upstream promoter and in the activation of the principal one (Malon & McClure, 1984). The binding of cAMP receptor protein seems to physically interfere with the entrance of RNA polymerase into the upstream promoter, It is possible that at the larval stage of *Drosophila* when the distal (adult) promoter of the *Adh* gene is inhibited by binding of a regulatory protein, if present, in *Drosophila*, the proximal (larval) promoter is expressed. However, it

seems difficult for the regulatory protein to activate the proximal promoter concomitantly in a manner similar to the cAMP receptor protein, because the proximal promoter is about 700 bp far away from the distal .

Here, I propose that such coordinated expression of two transcriptional units in a dual promoter gene is achieved by conformational alteration of chromatin topology. A particular topological conformation of chromatin would facilitate the establishment of a transcriptional complex by binding of regulatory factors at one particular promoter site but would preclude establishment at the other promoter site. Another topological conformation would achieve an alternate effect. Assuming that these genes are located within a topological domain of chromatin loop structure, such topological conformation could be easily altered by torsional stress. The two different conformations mentioned above must be alternatively formed in a topological domain by competing of secondary DNA structures for torsional energy as described previously (Chapter 4.2). I believe that such a regulatory mechanism is involved in the regulation by gene territoriality.

In fact, Emerman and Temin (1986) have recently found in their experiments on the "promoter suppression" in an infectious retrovirus vector that there is an inverse correlation between the DNase I sensitivity of the chromatin surrounding a promoter and the suppression of its use.

In contrast to the "promoter suppression", Kucherlapati and his colleagues have recently reported that genes introduced into cells by transfection can be concertedly expressed with their own promoters at the same time. Probably, the domain of this regulated expression is at least 20 kb long, the regulation may be associated with chromatin structure (Davies et al., 1981; Roginski et al., 1983). The apparent discrepancy between their and Emerman and Temin's results may be accounted for by the gene territoriality hypothesis since the relevant genes are on different strands. The genes that are linked together with a short intergenic distance between them in the same direction are predicted to be under territorial confrontation and consequently they may be alternatively expressed or both genes may become transcriptionally inactivated. On the other hand, the genes that are linked in the opposite direction are completely free from territorial effects. In this situation, both genes may be expressed at the same time. The latter case can partly explain the observation by Kucherlapati et al.

Thus when genes are clustered in a single domain, expression of those genes is coordinately regulated in various ways. The mechanism involving territorial effect operates only when they are on the same DNA strand. Whatever the case, such cooperative regulation within a domain must be achieved by the structural features of chromatin rather than by DNA sequences themselves alone. The results obtained in this thesis clearly indicate that

structural features can be dynamically altered by a long range effect through torsional stress.

Adams, J.M., Gerondakis, S., Webb, E., Corcoran, L.M. & Cory, S. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 1982-1986.

Adams, J.M. (1984). Presented at the 3rd Meeting of Aust. New Zealand Soc. Cell Biol., Melbourne.

Arnott, S., Chandrasekarn, R., Hall, I.H. & Puigjaner, L.C. (1983). Nucleic Acids Res. **11** 4141-4155.

ar-Rushdi, A., Nishikura, K., Erikson, J., Watt, R., Rovera, G. & Croce, C.M. (1983). Science **222**, 390-393.

Batterham, p., Lovett, J.A., Strarmer, W.T. & Sullivan, D.J. (1983). Dev. Biol. **96**, 346-354.

Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G. & Leder, P. (1983). Cell **34**, 770-789.

Beachy, P.A., Helfand, S.L. & Hogness, D.S. (1985). Nature **313**, 545-551.

Beard, P., Hughes, M., Nyfeler, K. & Moey, M. (1984). Eur. J. Biochem. **143**, 39-45.

Behe, M. & Felsenfeld, G. (1981). Proc. Nat. **78**, 1619-1623.

Bell, G.I., Gerhard, D.S., Fong, N.M., Sanchez-Pescador, R. & Ball, L.B. (1985). Proc. Natl. Acad. Sci. U.S.A. **82**, 6450-6454.

Bender, W., Akam, M., Karch, F., Beachy, P.A., Peifer, M., Spierer, P., Lewis, E.B. & Hogness, D.S. (1983). Science **221**, 23-29.

Benham, C.J. (1979). Proc. Natl. Acad. Sci. U.S.A. **76**, 3870-3874.

Benham, C.J. (1981). J. Mol. Biol. **150**, 43-68.

Benham, C.J. (1982). Biopolymers **21**, 679-696.

Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). Nucleic Acids Res. **8**, 127-182.

Benoist, C. & Chambon, P. (1981). Nature **290**, 304-310.

Benyajati, C. & Worcel, A. (1976). Cell **9**, 393-407.

Benyajati, C., Place, A.R., Wang, N., Pentz, E. & Sofer, W. (1982). Nucleic Acids Res. **10**, 7261-7272.

Benyajati, C., Spoerel, N., Haymerk, H. & Ashburner, M. (1983). Cell **33**, 125-133.

Benyajati, C. & Dray, J.F. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 1701-1705.

Bernard, O., Cory, S., Gerondakis, S., Webb, E. & Adams, J.M. (1983). EMBO J. **2**, 2375-2383.

Borgmeyer, U., Nowock, J. & Sippel, A.E. (1984). Nucleic Acids Res. **12**, 4295 - 4311.

Borras, T., Nickerson, J.M., Chepelinsky, A.B. & Piatigarsky, J. (1985). EMBO J. **4**, 445-452.

Breathnach, R. & Chambon, P. (1981). Ann. Rev. Biochem. **50**, 349-383.

Breidl, M., Harbers, K. & Jaenish, R. (1984). Cell **38**, 9-16.

Breslauer, K.J., Frank, R., Blocker, H. & Marky, L. (1986). Proc. Natl. Acad. Sci. U.S.A. **83**, 3746-3750.

Brutlag, D.C. Clayton, J. Friend, P. & Kedes, L.H. (1982). Nucleic Acids Res. **10**, 279-294.

Buongiorno-Nardelli, M., Micheli, G., Carri, M.T. & Marilley, M. (1982). Nature **298**, 100-102.

Burch, J.B.E. & Weintraub, H. (1983). Cell **33**, 65-76.

Calladine, C.R. (1982). J. Mol. Biol. **161**, 343-352.

Campisi, J., Gray, H.E., Pardee, A.B., Dean, M. & Sononshein, G.E. (1984). Cell, **36**, 241-247.

Canaani, E., Dreazen, O., Klar, A., Rechavi, G., Ram, D., Cohen, J.B. & Givol, D. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 7118-7122.

Canton, C.R. & Efstratiadis, A. (1984). Nucleic Acids Res. **12**, 8059-8072.

Cavalier-Smith, T. (1985). The Evolution Of Genome Size (John Wiley & Sons, London).

Chelm, B.K. & Geiduschek, E.P. (1979). Nucleic Acids Res. **7**, 1851-1867.

Chen, H.H., Behe, M.J. & Rau, D.C. (1984). Nucleic Acids Res. **12**, 2381-2389.

Cristophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V. & Vassart, G. (1985). Nucleic Acids Res. **13**, 5427-5744.

Cory, S., Gerondakis, S. & Adams, J.M. (1983). EMBO J. **2**, 697-703.

Courey, A.J. & Wang, J.C. (1983). Cell **33**, 817-829.

Croce, C.M., Erikson, J., ar-Rushdi, A., Aden, D. & Nishikura, K. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 3170-3174.

Davidson, E.H. (1976). Gene Activity in Early Development, 2nd Ed. (Academic Press, New York)

Dickerson, R.E. & Drew, H.R. (1981). J. Mol. Biol. **149**, 761-786.

Dierks, P., van Ooyen, A., Cochran, M.D., Dobkin, C., Reiser, J. & Weissman, C. (1983). Cell **32**, 695-706.

Dolan, M., Sugerman, B.J., Dodgson, J.B. & Engel, J.D. (1981). Cell **24**, 669-677.

Drew, H.R. & Dickerson, R.E. (1981) J. Mol. Biol. **151**, 535-536.

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn P.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S., Slighton, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulder, O.O. & Proudfoot, N.J. (1980). Cell **21**, 653-668.

Eissenberg, J.C., Cartwright, I.L., Thomas, G.H. & Elgin, S.C.R. (1985). Ann. Rev. Genet. **19**, 485-536.

Emerman,M. & Temin, H.M. (1986). Nucleic Acids Res. **14**, 9381-9396.

Emerson, B.M. & Felsenfeld, G. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 95-99.

Emerson, B.M., Lewis, C.D. & Felsenfeld, G. (1985). Cell **41**, 21-30.

Evans, T., Schon, E., Gora-Maslak, G., Patterson, J. & Efstratiadis, A. (1984). Nucleic Acids Res. **12**, 8043-8058.

Feuerstein, B.G., Pattabiraman, N. & Martin, L.J. (1986). Proc. Natl. Acad,. Sci. U.S.A. **83**, 5948-5952.

Fisher, J.A. & Maniatis, T. (1985). Nucleic Acids Res. **13**, 6899-6917.

Fisher, J.A. & Maniatis, T., (1986). EMBO J. **5**, 1275-1289.

Fleichmann, G., Pflugfelder, G., Steiner, E.K., Javaherian, K., Howard, G., Wang, J.C. & Elgin, S.C.R. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 6958-6962.

Fritton, H.P., Igo-Kemenes, T., Nowock, J., Strech-Jurk, U., Theisen, M. & Sippel, A.E. (1984). Nature **311**, 163-165.

Fujita, T., Ohno, S., Yasimitsu, H. & Taniguchi, T. (1985). Cell **41**, 489-496.

Gall[Bwitz, D., Donath, C. & Sander, C. (1983). Nature **306**, 704-707.

Gamper, H.B. & Hearst, J.E. (1982). Cell **29**, 81-90.

Garfinkel, M.D., Pruitt, R.E. & Meyerowitz, E.M. (1983). J. Mol. Biol. **168**, 765-789.

Gasule, L.C. & Schellman, J.A. (1976). Nature **239**, 333-335.

Gellert, M., Mizuuchi, K., O'Dea, M.H. & Nash, H.A. (1976). Proc. Natl. Acad. Sci. U.S.A. **73**, 3872-3876.

Germond, J.E., Hirt, B., Gross-Bellard, M., Chambon, P. (1975). Proc. Natl. Acad. Sci. U.S.A. **72**, 1843-1847.

Gilmour, D.S., Plugfelder, G., Wang, J.C. & Lis. J.T. (1986). Cell **44**, 401-407.

Glikin, G.C., Gargiulo, G., Rena-Descalzi, L. & Worcel, A. (1983). Nature **303**, 770-774.

Glikin, G.C., Ruberti, I. & Worcel, A. (1984). Cell **37**, 33-41.

Glisin, V. Crvenjakow, R., Byus, C. (1974). Biochemistry **13**, 2633-2637.

Goding, C.R. & Russell, W.C. (1983). Nucleic Acids Res. **11**, 21-36.

Goldberg, D.A. (1980). Proc Natl. Acad. Sci. U.S.A. **77**, 5794-5798.

Goldberg, D.A., Posakony, J.W. & Maniatis, T. (1983). Cell **34**, 59-73.

Gottesfeld, J. & Bloomer, L.S. (1982). Cell **28**, 781-791.

Graham, D.E., Neufeld, B.R., Davidson, E.H. & Britten, R.J. (1974). Cell **1**, 127-137.

Greaves, D.R., Patient, R.K. & Lilley, D.M.J. (1985). J. Mol. Biol. **185**, 461-478.

Green, M.H., Buss, J. & Gariglio, P. (1975). Eur. J. Biochem. **53**, 217-225.

Groudine, M. & Weintraub, H. (1981). Cell **24**, 393-401.

Groudine, M. & Weintraub, H. (1982). Cell **30**, 131.

Gruss, P. (1984). DNA **3**, 1-5.

Gurdon, J.B. & Wickens, M.P. (1983). Methods Enzymol. **101**, 370-386.

Han, S., Udvardy, A. & Schedl, P. (1984). J. Mol. Biol. **179**, 469-496.

Hanas, J.S., Bogenhagen, D.F. & Wu, C.W. (1984). Nucleic Acids Res. **12**, 1265-1276.

Haniford, D.B. & Pulleybank, D.E. (1985). Nucleic Acids Res. **13**, 4343-4363.

Harland, R.M. Weintraub, H. & McKnight, S.L. (1983). Nature **302**, 38-43.

Hattori, M. & Sakaki, Y. (1986). Analyt. Biochem. **152**, 232-238.

Hayday, A.C., Gillies, S.D., Saito, H., Wood, C., Wiman, K., Hayward, W.S. & Tonegawa, S. (1984). Nature **307**, 334-340.

Hayward, W.S., Neel, B.G. & Astrin, S.M. (1981). Nature **290**, 475.

Heberlein, U. England, B. & Tjian, R. (1985). Cell **41**, 965-977.

Hentschel, C.C. (1982). Nature **295**, 714-716.

Hochschild, A.H. & Ptashne, M. (1986). Cell **44**, 681-687.

Hofer, E. & Darnell, J.E. (1981). Cell **23**, 385-393.

Hollis, G.F., Mitchell, K.F. Battey, J., Potter, H., Taub, R., Lenoir, G.M. & Leder, P. (1984). Nature **307**, 752-755.

Howkins, J.W., Nickerson, J.M., Sullivan, A.M. & Piatigorsky, J. (1984). J. Biol. Chem. **259**, 9821-9825.

Htun, H., Lund, E. & Dahlberg, J.E. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 7288-7292.

Iatrou, K. & Tsitilon, S.G. (1983). EMBO J. **2**, 1431-1440.

Ingram, R.S., Scott, R.W. & Tilghman, S.M. (1981). Proc. Natl. Acad. Sci. U.S.A. **78**, 4694-4698.

Jaenish, R., Jahjner, D., Nobis, P. Simon, I. Lohler, J. Harbers, K. & Grotkopp, D. (1981). Cell **24**, 519-529.

Jaenish, R., Harbers, K., Schneike, A., Lohler, J., Chumakov, I., Jahner, D., Grotkopp, D. & Hoffman, E. (1983). Cell **32**, 209-216.

Jaenish, R. & Soriano, P. (1986). Proc. Aust. Biochem. Soc. **18**, 528.

Jakobovits, E.B., Bratosin, S. & Aloni, Y. (1980). Nature **285**, 263-265.

Jantzen, K., Fritton, M.P. & Igo-Kemenes, T. (1986). Nucleic Acids Res. **14**, 6085-6099.

Jeffreys, A.J. (1982). In *Genome Evolution* (Dover, G.A. & Flavell, R.B. eds) p. 157. Academic Press, London.

Jensch, F. & Kemper, B. (1986). EMBO J. **5**, 181-189.

Jongstra, J., Reudelhuber, T.L., Oudet, P., Benoist, C., Chae, C.B. , Jetlsch, J.M., Mathis, D.J. & Chambon, P. (1984). Nature **307**, 708-714.

Jordano, J. & Perucho, M. (1986). Nucleic Acids Res. **14**, 7361-7378.

Kaplan, B.B., Bernstein, S.L. & Gioio, A.E. (1979). Biochem. J. **183**, 181-184.

Karathanasis, S. (1985). Proc. Natl. Acad. Sci. U.S.A. **82**, 6374-6378.

Kedes, L.H. (1979). Ann. Rev. Biochem. **48**, 837-870.

Keene, M.A. & Elgin, S.C.R. (1981). Cell **27**, 57-64.

Keene, M.A. & Elgin, S.C.R. (1984). Cell **36**, 121-129.

Kelleher III, R.J., Ellison, M.J., Ho, P.S. & Rich, A. (1986). Proc. Natl. Acad. Sci. U.S.A. **83**, 6342-6346.

Kelly, K., Cochrane, B.H., Shles, C.D. & Leder, P. (1983). Cell **35**, 603-610.

Kilpatrick, M.W., Wei, C.F., Gray, H.B. Jr. & Wells, R.D. (1983). Nucleic Acids Res. **11**, 3811-3822.

Kilpatrick, W.M., Kang, A.D., Engler, J.A. & Wells, R. (1986). J. Biol. Chem. **261**, 11350-11354.

Kioussis, D., Eiferman, F., van de Rijis, P., Gorin, M.B. Ingram, R.S. & Tilghman, S.M. (1981). J. Biolo. Chem. **256**, 1960-1967.

Klein, G. (1983). Cell **32**, 311-315.

Kohwi-Shigematsu, T., Gelinas, R. & Weintraub, H. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 4389-4393.

Kohwi-Shigematsu, T. & Kohwi, Y. (1985). Cell **43**, 199-206.

Kowalczykowski, S.C., Bear, D.G. & von Hippel, P.H. (1981). The Enzymes **14**, 373-444.

Kowalski, D. & Sanford, J.P. (1982). J. Biol. Chem. **257**, 7820-7825.

Kreitman, M. (1983). Nature **304**, 412-417.

Kretsovali, A., Marcaud, L., Moreau, J. & Scherrer, K. (1986). Mol. Gen. Genet. **203**, 193-201.

Kuroiwa, A., Kloter, U., Baumgartner, P. & Gehring, W.J. (1985). EMBO J. **4**, 3757-3764.

Landes, G.M., Villeponteau, B., Priby, T.M. & Martinson, H.G. (1982). J. Biol. Chem. **257**, 11008-11014.

Larsen, A. & Weintraub, H. (1982). Cell **29**, 609-622.

Leskey, R.A., Mills, A.D. & Morris, N.R. (1977). Cell **10**, 237-243.

Lawn, R.M., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980). Cell **21**, 647-651.

Lawson, G.M., Tsai, M.-J., O'Malley, B.W. (1980). Biochemistry **19**, 4403-4411.

Lawson, G.M. Knoll, B.J., March, C.J., Woo, S.L.C., Tsai, M.-J. & O'Malley, B.J. (1982). J. Biol. Chem. **257**, 1501-1507.

Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W.,Potter, H., Stewart, T. & Taub, R. (1983). Science **222**, 766-771.

Leff, S.E., Rosenfeld, M.G. & Evans,R.M. (1986). Ann. Rev. of Biochem. **55**, 1091-1117.

Lee, J.S., Johnson, D.A. & Morgan, A.R. (1979). Nucleic Acids Res. **6**, 3073-3091.

Lima-de-Faria, A. (1983). *Molecular evolution and organization of the chromosomes*. Elsevier, Amsterdam.

Lilley, D.M.J. (1980). Proc. Natl. Acad. Sci. U.S.A. **77**, 6468-6472.

Lilley, D.M.J. & Hollam, L.R. (1984). J. Mol. Biol. **180**, 179-200.

Lilley, D.M.J. & Houghton, M. (1979). Nucleic Acids Res. **6**, 507-523.

Loff, S.E., Rosenfeld, M.G. & Evans, R.M. (1986). Ann. Rev. Biochem. **55**, 1091-1117.

Lompre, A.M., Nadal-Ginard, B. & Mahdavi, V. (1984). J. Biol. Chem. **259**, 6437-6446.

Loomis, W.F. & Gilpin, M.E. (1986). Proc. Natl. Acad. Sci U.S.A. **83**, 2143-2147.

Luchnik, A.N., Bakayev, V.V., Zbarsky, I.B. & Georgiev, G.P. (1982). EMBO J. **1**, 1353-1359.

Luchnik, A.N., Bakayev, V.V., Yugai, A.A., Zbarsky, I.B. & Georgiev, G.P. (1985). Nucleic Acids Res. **13**, 1135-1149.

Mace, H.A., Pelham, R.B. & Travers, A.A. (1983). Nature **304**, 555-557.

Mahdavi, V., Chambers, A.P. & Nadal-Ginard, B. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 2626-2630.

Malan, T.P. & McClure, W.R. (1984). Cell **39**, 173-180.

Maniatis, T., Fritsch, E.F. & Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory, New York).

Manley, J.L., Fire, A., Cano, A., Sharp, P.A. & Gefter, M.L. (1980). Proc. Natl. Acad. Sci. U.S.A. **77**, 3855-3859.

Manley, J.L., Fire, A., Samuels, M. & Sharp, P.A. (1983). Methods Enzymol. **101**, 568-581.

Marcu, K.B., Harris, L.J., Stanton, L.W., Erikson, J., Watt, R. & Croce, C.M. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 519-523.

Margot, J.B. & Hardison, R.C. (1985). J. Mol. Biol. **184**, 195-210.

Mason, P.J., Torok, I., Kiss, I., Karch, F. & Udvardy, A. (1982). J. Mol. Biol. **156**, 21-35.

Maxam, A.M. & Gilbert, W. (1980). Methods Enzymol. **65**, 499-560.

McCready, S.J., Godwin, J., Mason, D.W., BrazellI.A. & Cook, P.R. (1980). J. Cell. Sci. **46**, 675-691.

McGhee, J.D. & Felsenfeld, G. (1980). Ann. Rev. Biochem. **49**, 1115-1156.

McGhee, J.D., Wood, W.I., Dolan, M., Engel, J.D. & Felsenfeld, G. (1984). Cell **27**, 45-55.

McGinnis, W., Shermoen, A.W., Heemskerk, J. & Beckendorf, S.K. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 1063-1067.

McKeon, C., Schmidt, A. & de Crombrugghe, B. (1984). J. Biol. Chem. **259**, 6636-6640.

Messing, J. (1983). Methods Enzymol. **101**, 20-89.

Miller, T.J. & Mertz, J.E. (1982). Mol. Cell. Biol. **2**, 1595-1607.

Mirkovitch, J., Spierer, P. & Laemmli, U.K. (1986). J. Mol. Biol. **190**, 255-258.

Mizuuchi, K., Kemper, B., Hays, J. & Weisberg, R.A. (1982). Cell **29**, 357-365.

Modolell, J., Bender, W. & Meselson, M. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 1678-1682.

Moreau, J., Matyash-Smirniaguina, L. & Scherrer, K. (1981). Proc. Natl. Acad. Sci. U.S.A. **78**, 1341-1345.

Moreau, J., Marcaud, L., Maschat, F., Kejzlarova-Lepesant, J., Lepesant, J.-A., Scherrer, K. (1982). Nature **295**, 260-262.

Muller, R., Slamon, D.J., Tremblay, J.M., Cline, M.J. & Verma, I.M. (1982). Nature **299**, 640-643.

Naora, H. & Deacon, N.J. (1982a). Differentiation **21**, 1-6.

Naora, H. & Deacon, N.J. (1982b). J. Theor. Biol. **95**, 601-606.

Naora, H., Deacon, N.J. & Braithwaite, A.W. (1983). J. Theor. Biol. **105**, 715-721.

Naora, H. (1986). Biol. Forum **79**, 345-371.

Nickol, J.M. & Felsenfeld, G. (1983). Cell **35**, 467-477.

Nickol, J.M. & Martin, R.G. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 4669-4683.

Nishikura, K., ar-Rushdi, A., Erikson, J., Watt, R., Rovera, G. & Croce, C.M. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 4822-4826.

Nusse, R., van Ooyen, A., Cox, D., Fung, Y.K.T. & Varmus, H. (1984). Nature **307**, 131-136.

Nussinov, R., Shapiro, B., Lipkin, L.E. & Maizel Jr, J.V. (1984). J. Mol. Biol. **177**, 591-607.

Ohno, S. (1970). *Evolution by gene duplication* (Springer-Verlag, New York).

Oskarsson, M., McClements, W.L., Blair, D.G., Maizel, J.V. & Vande Woude, G.F. (1980). Science **207**, 1222-1224.

Panayotatos, N. & Wells, R.P. (1981). Nature **289**, 466-470.

Parkhurst, S.M. & Corces, V.G. (1985). Cell **41**, 429-437.

Parkhurst, S.M. & Corces, V.G. (1986). Mol. Cell. Biol. **6**, 47-53.

Patient, R.K., Elkington, J.A., Kay, R.M. & Williams, J.G. (1980). Cell **21**, 565-573.

Paulson, J.R. & Laemmli, U.K. (1977). Cell **12**, 817-828.

Payne, G.S., Bishop, J.M. & Varmus, H. (1982). Nature **295**, 209-214.

Peck, L.J. & Wang, J.C. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 6206-6210.

Peck, L.J., Nordheim, A., Rich, A. & Wang, J.C. (1982). Proc Natl. Acad. Sci. U.S.A. **79**, 4560-4564.

Peifer, M. & Bender, W. (1986). EMBO J. **5**, 2293-2303.

Pelham, H.R.B. (1982). Cell **30**, 517-528.

Perry, R.P. (1983). Cell **33**, 647-649.

Pienta, K.J. & Coffey, D.S. (1984). J. Cell Sci. Suppl. **1**, 123-135.

Pratt, J.M. (1984). In *Transcritpion and translation: a practical approach* pp. 179-209. (IRL Press, Oxford).

Probst, E., Kressmann, A. & Birnstiol, M.L. (1979). J. Mol. Biol. **135**, 709-732.

Proudfoot, N.J. & Brownlee, G.G. (1976). Nature **263**, 211-214.

Proudfoot, N.J. (1986). Nature **322**, 562-565.

Pulleybank, D.E., Haniford, D.B. & Morgan, A.R. (1985). Cell **42**, 271-280.

Rabbitts, T.H., Hamlyn, P.H. & Baer, R. (1983). Nature **306**, 760-765.

Reynolds, W.F. & Gottesfeld, J.M. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 1862-1866.

Robertson, M. (1983). Nature **306**, 733-736.

Ryoji, M. & Worcel, A. (1984). Cell **37**, 21-32.

Saenger, W. (1983). Principle of Nucleic Acid Structure. (Springer-Verlag, New York).

Saragosti, S., Moyne, G. & Yaniv, M. (1980). Cell **20**, 65-73.

Savakis, C., Ashburner, M. & Willis, J.H. (1986). Dev. Biol. **114**, 194-207.

Schneike, A., Harbers, K. & Jaenish, R. (1983). Nature **304**, 315-320.

Scholten, P.M. & Norhei, A. (1986). Nucleic Acids Res. **14**, 3981-3993.

Schon, E., Evans, T., Welsh, J. & Efstratiadis, A. (1983). Cell **35**, 837-848.

Selleck, S.B., Elgin, S.C.R. & Cartwright, I.L. (1984). J. Mol. Biol. **178**, 17-33.

Sergeant, A., Bohmann, D., Zentgraf, H., Weiher, H., Keller, W. (1984). J. Mol. Biol. **180**, 577-600.

Shaw, P., Sordat, B. & Schibler, U. (1985). Cell **40**, 907-912.

Sheflin, L.G. & Kowalsk, D. (1984). Nucleic Acids Res. **12**, 7087-7104.

Shermoen, A.W. & Beckendorf, S.K. (1982). Cell **29**, 601-607.

Shure, M., Pulleybank, D.E. & Vinograd, J. (1977). Nucleic Acids Res. **4**, 1183-1205.

Sigal, N., Delius, H. Kornberg, T., Gefter, M.L. & Alberts, B. (1972). Proc. Natl. Acad. Sci. U.S.A. **69**, 3537-3541.

Sinden, R., Carlson, J.O. & Pettijohn, D.E. (1980). Cell **21**, 773-783.

Singleton, C.K. & Wells, R.D. (1982a). Analyt. Biochem. **122**, 253-257.

Singleton, C.K. & Wells, R.D. (1982b). J. Biol. Chem. **257**, 6292-6295.

Singleton, C.K., Klysik, J., Stirdivant, S.M. & Wells, R.D. (1982). Nature **299**, 312-316.

Southern, E.M. (1985). J. Mol. Biol. **98**, 503-517.

Spritz, R.A., DeRiel, J.K., Forget, B.G. & Weissman, S.M. (1980). Cell **21**, 639-646.

Srivenugopal, K.S. & Morris, D.R. (1985). **24**, 4766-4771.

Stalder, J., Groudine, M., Dodgson, J.B., Engel, J.P. & Weintraub, H. (1980). Cell **19**, 973-980.

Stalder, J., Larsen, A., Engel, J.D. Dolan, M., Groudine, M. & Weintraub, H. (1980). Cell **20**, 451-460.

Stanton, L.W. Yang, J.-Q., Eckhart, L.A., Harris, L.J., Birshtein, B.K. & Marcu, K.B. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 829-833.

Stephenson, E.C., Erba, H.P. & Gall, J. G. (1981). Nucleic Acids Res. **9**, 2281-2295.

Stonington, O.G. & Pettijohn, D.E. (1971). Proc. Natl. Acad. Sci. **68**, 6-9.

Taub, R., Moulding, C., Battey, J., Murphy, W., Vasicek, T., Lenoir, G.M. & Leder, P. (1984). Cell **36**, 339-348.

Tolunay, H.E., Yang, L., Anderson, W.F. & Safer, B. (1984). Proc. Natl. Acad. Sci. U.S.A. **81**, 5916-5920.

Trifonov, E.N. (1980). Nucleic Acids Res. **8**, 4041-4053.

Trifonov, E.N. & Sussman, J.L. (1980). Proc. Natl. Acad. Sci. U.S.A. **77**, 3816-3820.

Ulanovsky, L., Bodner, M., Trifonov, E.N. & Choder, M. (1986). Proc. Natl. Acad. Sci. U.S.A. **83**, 862-866.

Ursprung, H., Sofer, W. & Burroughs, N. (1970). Wilhelm Roux's Arch. Dev. Biol. **164**, 201-208.

Villeponteau, B., Lundell, M. & Martinson, M. (1984). Cell **39**, 469-478.

Vogelstein, B., Pardoll, D.M. & Coffey, D.S. (1980). Cell **22**, 79-85.

Vologodskii, A.V., Lukashin, A.V., Anshelevich, V.V. & Frank-Kamenetskii, M.D. (1979). Nucleic Acids Res. **6**, 967-982.

Wang, J.C. (1974). J. Mol. Biol. **87**, 797-816.

Wang, J.C. & Hogan, M. (1985). J. Biol. Chem. **260**, 8194-8202.

Wang, J.C., Peck, L.J. & Becherer, K. (1983). Cold Spring Harbor Symp. Quant. Biol. **47**, 85-91.

Webster, G. & Goodwin, B.C. (1984). Riv. Biol. **77**, 503-

Weintraub, H. & Groudine, M. (1976). Science **93**, 848-853.

Weintraub, H. (1983). Cell **32**, 1191-1203.

Weintraub, H. (1985). Cell **42**, 705-711.

Wilbur, W.J. & Lipman, D.J. (1983). Proc. Natl. Acad. Sci. U.S.A. **80**, 726-730.

Wu, C., (1980). Nature **286**, 854-860.

Wu, C., (1984a). Nature **309**, 229-234.

Wu, C., (1984b). Nature **311**, 81-84.

Yang, R.C.-A., Lis, J. & Wu, R. (1979). Methods Enzymol. **68**, 176-182.

Yaniv, M. (1984). Biol. Cell **50**, 203-266.

Young, R.A., Hagenbuchle, O. & Schibler, U. (1981). Cell **23**, 451-458.

Yunis, J. (1983). Science **221**, 227-236.